



# Partitionnement des images hyperspectrales de grande dimension spatiale par propagation d'affinité

Mariem Soltani

## ► To cite this version:

Mariem Soltani. Partitionnement des images hyperspectrales de grande dimension spatiale par propagation d'affinité. Traitement du signal et de l'image [eess.SP]. Université de Rennes, 2014. Français. NNT : 2014REN1S099 . tel-01134984

**HAL Id: tel-01134984**

**<https://theses.hal.science/tel-01134984>**

Submitted on 26 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE / UNIVERSITÉ DE RENNES 1**

*sous le sceau de l'Université Européenne de Bretagne*

pour le grade de

**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Traitement du Signal et Télécommunications*

**Ecole doctorale MATISSE** (Mathématiques, Télécommunications,  
Informatique, Signal, Systèmes, Electronique)

présentée par

**Mariam Soltani**

Préparée à l'unité de recherche 6164 IETR  
(Institut d'Électronique et de Télécommunications de Rennes)

Équipe : Traitement des Signaux & Images Multicomposantes et Multimodales

---

**Intitulé de la thèse :**

**Partitionnement  
des images  
hyperspectrales de  
grande dimension  
spatiale par  
propagation  
d'affinité**

**Thèse soutenue à l'Université de Rennes 1  
le 17 décembre 2014**

devant le jury composé de :

**M. Azeddine BEGHDADI**

Professeur à l'Institut Galilée - Université Paris 13 /  
*rapporteur*

**M. Didier COQUIN**

Professeur à Polytech Annecy / *rapporteur*

**M. Franck MARZANI**

Professeur à l'Université de Bourgogne / *examineur*

**M. Claude CARIOU**

Maître de conférences à l'Université de Rennes 1 /  
*examineur*

**M. Kacem CHEHDI**

Professeur à l'Université de Rennes 1 / *directeur de thèse*



# Notation

$I$  : image à partitionner

$X = \{x_1, \dots, x_i, \dots, x_N\}$  : ensemble de  $N$  individus (pixels) à classer

$\ell$  : un niveau de gris

$\ell_i$  : niveau de gris d'un pixel  $x_i$

$N_\ell$  : nombre de niveaux de gris

$N_a$  : nombre d'attributs représentant un individu  $x_i$

$N_c$  : nombre de classes d'une image  $I$

$a_i = (a_{i1}, a_{i2}, \dots, a_{iN_a})$  : vecteur d'attributs (ou spectre) de dimension  $N_a$  représentant l'individu  $x_i$

$A_j$  :  $[a_{1j}, a_{2j}, \dots, a_{Nj}]^T$  attribut  $j$  représentant l'ensemble des  $N$  individus.

$P^r = \{P^r_1, P^r_2, \dots, P^r_{N_c}\}$  : ensemble des prototypes des  $N_c$  classes.

$C_i$  : classe  $i$  de l'image  $I$

$N_{C_i}$  : nombre d'individus de la classe  $C_i$

$g_i$  : centre de gravité de la classe  $C_i$

$\sigma$  : écart-type

$D_j(x_i)$  : degré d'appartenance de  $x_i$  à la classe  $C_j$

$\mu$  : moyenne

$d(., .)$  : distance

$V_{xi}$  : voisinage d'un pixel  $x_i$

$S$  : matrice de similarité

$R$  : matrice de responsabilité

$A$  : matrice de disponibilité

$p$  : paramètre de préférence

$\lambda_i$  : longueur d'onde de la bande  $i$

$R(\lambda_i)$  : mesure de réflectance à la longueur d'onde  $\lambda_i$

## Résumé

Les images hyperspectrales suscitent un intérêt croissant depuis une quinzaine d'années. Elles fournissent une information plus détaillée des propriétés spectrales d'une scène et permettent une discrimination plus précise des objets que les images couleur ou même les images multispectrales. Bien que les potentialités de la technologie hyperspectrale apparaissent relativement grandes, l'analyse et le traitement de ces importantes quantités de données restent une tâche difficile et leur exploitation présente aujourd'hui un défi en termes d'interprétation.

Les travaux de cette thèse s'inscrivent dans le cadre de la réduction et du partitionnement des images hyperspectrales de grande dimension spatiale. L'approche proposée se compose essentiellement de deux étapes : extraction d'attributs et classification des pixels de l'image. Il existe dans la littérature différents travaux consacrés au développement des méthodes statistiques de caractérisation des pixels allant des moments d'ordre 1, 2 et 3, aux attributs calculés à partir des matrices de cooccurrences et de transformée en ondelettes. Une nouvelle approche d'extraction d'attributs à partir des matrices de tri-occurrences définies sur des voisinages cubiques est proposée. Ces attributs tiennent compte conjointement de l'information spatiale et spectrale des données. Une étude comparative a été menée afin de tester le pouvoir discriminant de ces nouveaux attributs par rapport aux attributs classiques. Les attributs proposés montrent un large écart discriminant par rapport à ces derniers et par rapport aux signatures spectrales originales.

Pour pouvoir exploiter ces données, l'opération de classification constitue une étape incontournable. Cette opération peut être menée de manière supervisée, semi-supervisée ou non supervisée. La classification non supervisée présente plusieurs avantages par rapport aux autres approches, tels que la non nécessité pour l'utilisateur de préciser à l'avance les classes à discriminer et les échantillons d'apprentissage (relevés de terrain) dont l'accès pour certaines applications est très difficile. C'est pourquoi, nous nous intéressons ici au partitionnement des images par une approche de classification non supervisée et non paramétrique. Le caractère non paramétrique permet d'éviter les erreurs introduites lors de l'estimation des paramètres des modèles qui peuvent fausser la classification et plus particulièrement dans le cas des données de grande dimension. Une étude comparative des principaux classifieurs semi-supervisés (connaissance du nombre de classes) et non supervisés non paramétriques (C-moyennes, FCM, ISODATA, AP, ...), a montré la supériorité de la méthode de propagation d'affinité (AP). Mais malgré un meilleur taux de classification, cette méthode présente deux inconvénients majeurs. D'une part, la version complètement non supervisée donne une surestimation du nombre de classes. D'autre part, lorsqu'il s'agit de partitionner des images hyperspectrales de grande taille spatiale, sa complexité de calcul est quadratique. Par conséquent, son application sur ces types de données reste impossible. Nous avons proposé une approche qui apporte des solutions à ces deux problèmes. Elle consiste tout d'abord à réduire le nombre d'individus à classer avant l'application de l'AP en introduisant une étape de pré-classification qui agrège automatiquement les pixels à très forte similarité. L'image hyperspectrale est divisée en blocs, puis l'étape de réduction est appliquée indépendamment sur chacun des blocs. L'AP est ensuite appliquée sur un nouvel ensemble de pixels constitué cette fois-ci des représentants de chaque groupe formé et des individus non agrégés. Pour estimer le nombre de classes, la méthode AP utilise de manière implicite un paramètre de préférence  $p$  dont la valeur initiale correspond à la médiane des valeurs de la matrice de similarité. Cette valeur conduisant souvent à une sur-segmentation des images, nous avons introduit une étape permettant de l'optimiser en maximisant un critère lié à la variance interclasse. L'approche proposée a été testée sur des images synthétiques, monocomposantes et multicomposantes et a montré pour l'ensemble des tests réalisés, une discrimination cohérente des différentes classes dans les images. Elle a été également appliquée et comparée sur des images hyperspectrales de grande taille spatiale ( $1000 \times 1000$  pixels  $\times$  62 bandes avec une résolution spatiale au sol de 0.5m) dans le cadre d'une application réelle pour la détection des plantes invasives et non invasives. L'approche non supervisée développée a donné sur cette grande image un taux moyen de bonne classification de 97.19% contre 76.64% pour FCM, 65.03% pour C-moyenne et 62.81% pour ISODATA. Deux autres applications concernant l'identification du couvert paysager sont également traitées.

**Mots clés :** partitionnement, réduction, classification, non supervisée, non paramétrique, propagation d'affinité, imagerie hyperspectrale, grande taille spatiale, validation, détection, plantes invasives.

## Abstract

The interest in hyperspectral image data has been constantly increasing during the last years. Indeed, hyperspectral images provide more detailed information about the spectral properties of a scene and allow a more precise discrimination of objects than traditional color images or even multispectral images. High spatial and spectral resolutions of hyperspectral images enable to precisely characterize the information pixel content. Though the potentialities of hyperspectral technology appear to be relatively wide, the analysis and the treatment of these data remain complex. In fact, exploiting such large data sets presents a great challenge.

In this thesis, we are mainly interested in the reduction and partitioning of hyperspectral images of high spatial dimension. The proposed approach consists essentially of two steps: features extraction and classification of pixels of an image. A new approach for features extraction based on spatial and spectral tri-occurrences matrices defined on cubic neighborhoods is proposed. A comparative study shows the discrimination power of these new features over conventional ones as well as spectral signatures. Concerning the classification step, we are mainly interested in this thesis to the unsupervised and non-parametric classification approach because it has several advantages: no *a priori* knowledge, image partitioning for any application domain, and adaptability to the image information content. A comparative study of the most well-known semi-supervised (knowledge of number of classes) and unsupervised non-parametric methods (K-means, FCM, ISODATA, AP) showed the superiority of affinity propagation (AP). Despite its high correct classification rate, affinity propagation has two major drawbacks. Firstly, the number of classes is over-estimated when the preference parameter  $p$  value is initialized as the median value of the similarity matrix. Secondly, the partitioning of large size hyperspectral images is hampered by its quadratic computational complexity. Therefore, its application to this data type remains impossible. To overcome these two drawbacks, we propose an approach which consists of reducing the number of pixels to be classified before the application of AP by automatically grouping data points with high similarity. We also introduce a step to optimize the preference parameter value by maximizing a criterion related to the interclass variance, in order to correctly estimate the number of classes.

The proposed approach was successfully applied on synthetic images, monocomponent and multicomponent and showed a consistent discrimination of obtained classes. It was also successfully applied and compared on hyperspectral images of high spatial dimension ( $1000 \times 1000$  pixels  $\times$  62 bands) in the context of a real application for the detection of invasive and non-invasive vegetation species. Two other results concerning the application of the identification of the landscape covered are also presented.

**Keywords:** partitioning, reduction, classification, unsupervised, non-parametric, affinity propagation, image, hyperspectral, large spatial size, validation, detection, invasive vegetation, landscape identification.

## *Remerciements*

Il me sera très difficile de remercier tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme.

Je voudrais tout d'abord remercier grandement mon directeur de thèse, Kacem CHEHDI, pour son aide. Je suis ravie d'avoir travaillé en sa compagnie car outre son appui scientifique, il a toujours été là pour me soutenir et me conseiller au cours de l'élaboration de cette thèse. Je tiens à remercier particulièrement Claude CARIOU pour toutes nos discussions et ses conseils qui m'ont accompagné tout au long de mes années de thèse.

Azeddine BEGHDADI et Didier COQUIN m'ont fait l'honneur d'être rapporteurs de ma thèse. Leurs remarques m'ont permis d'envisager mon travail sous un autre angle. Pour tout cela je les remercie. Je tiens à remercier Franck MARZANI pour l'honneur qu'il me fait d'être mon président de jury.

Au cours de ces années j'ai fait partie de l'équipe «Traitement des Signaux & Images Multicomposantes et Multimodales» de l'IETR. Les discussions que j'ai pu avoir durant les réunions d'équipe ou en dehors avec Josias LEVRE et Akar TAHER m'ont beaucoup apporté.

Je tiens à remercier ma famille et mes amis pour leur soutien.

Mes derniers remerciements vont à RAMI qui a tout fait pour m'aider, qui m'a soutenu et surtout supporté dans tout ce que j'ai entrepris.

Encore un grand merci à tous pour m'avoir conduit à ce jour mémorable

## Table des matières

Introduction Générale .....	7
Partie I.....	11
État de l'art.....	11
Chapitre 1 .....	12
Description du contexte et problématique .....	12
1.1. Contexte général .....	12
1.1.1. Imagerie hyperspectrale. ....	12
1.1.2. Domaines applicatifs. ....	13
1.2. Problèmes liés à la classification des données en grande dimension .....	14
1.2.1. Choix des attributs - Information spatiale et spectrale .....	14
1.2.2. Choix de la méthode de classification .....	15
1.3. Conclusion .....	16
Chapitre 2 .....	18
Caractérisation des individus.....	18
2.1. Introduction .....	18
2.2. Attributs spatiaux .....	19
2.2.1. Les moments statistiques.....	19
2.2.2. Les attributs de la matrice de cooccurrences.....	21
2.2.3. Les attributs des longueurs de plages.....	22
2.2.4. Discussion .....	22
2.3. Attributs issus de la transformée en ondelettes.....	23
2.4. Les statistiques d'ordre supérieur.....	27
2.5. Attributs spectraux.....	27
2.6. Attributs spatiaux et spectraux .....	28
2.7. Conclusion .....	32
Chapitre 3 .....	33
Classification semi-supervisée et non supervisée.....	33
3.1. Introduction .....	33
3.2. La classification.....	35
3.2.1. Méthodes de classification semi-supervisées .....	35
3.2.1.1. La famille des C-Moyennes .....	36
3.2.1.2 La méthode ISODATA .....	38
3.2.1.3 La carte auto-organisatrice .....	40
3.2.1.4 L'immunité artificielle.....	42
3.2.2. Méthodes de classification non supervisée .....	45
3.2.2.1 Les méthodes génétiques .....	46
3.2.2.2 FCM non supervisé.....	48
3.2.2.3 La méthode MLBG (Modified LBG).....	49
3.2.2.4 La propagation d'affinité .....	50
3.3. Analyse de l'influence des attributs et du choix de la métrique sur les résultats de la classification par AP .....	58
3.4. Discussion .....	59
Chapitre 4 .....	64
Critères d'évaluation d'une méthode de classification .....	64
4.1. Introduction.....	64
4.2. Critères d'évaluation non supervisés. ....	65
4.3. Analyse des indices d'évaluation associés à l'AP. ....	72
Conclusion de la Partie I.....	77
Partie II .....	79
Approche de partitionnement développée et évaluation.....	79
Chapitre 5 : Approche proposée.....	80



<b>5.1. Introduction .....</b>	<b>80</b>
<b>5.2. Extraction d'attributs de tri-occurrences.....</b>	<b>81</b>
5.2.1. Attributs de tri-occurrences spatiaux.....	82
5.2.2. Validation .....	85
5.2.3. Conclusion.....	88
5.2.4. Attributs de tri-occurrences spatiaux et spectraux.....	89
5.2.5. Validation .....	90
5.2.6. Conclusion.....	93
<b>5.3. Réduction de dimension et classification non supervisée par propagation d'affinité.....</b>	<b>93</b>
5.3.1. Réduction du nombre d'individus à classer .....	93
5.3.2. Validation de l'étape de réduction.....	95
5.3.3 Classification avec optimisation de l'estimation du nombre de classes.....	101
<b>5.4. Conclusion .....</b>	<b>103</b>
<b>Chapitre 6 .....</b>	<b>104</b>
<b>Application à l'identification de plantes invasives et du couvert paysager.....</b>	<b>104</b>
<b>6.1 Identification de plantes invasives.....</b>	<b>104</b>
6.1.1 Présentation des données.....	104
6.1.2 Résultats d'identification.....	105
<b>6.2 Identification du couvert paysager.....</b>	<b>112</b>
6.2.1 Détection des pins .....	112
6.2.1 Identification des structures au sol .....	115
<b>6.4. Conclusion .....</b>	<b>116</b>
<b>Conclusion générale .....</b>	<b>117</b>
<b>Annexe A : .....</b>	<b>119</b>
<b>Attributs spatiaux pour images monocomposantes .....</b>	<b>119</b>
<b>Annexe B : .....</b>	<b>123</b>
<b>Critères d'évaluation comparés.....</b>	<b>123</b>
<b>Bibliographie .....</b>	<b>124</b>
<b>Liste des tableaux .....</b>	<b>136</b>
<b>Liste des figures .....</b>	<b>136</b>

## Introduction Générale

L'intérêt pour les données issues de capteurs d'imagerie hyperspectraux est croissant depuis quelques années. En effet, les images hyperspectrales fournissent une information plus détaillée des propriétés spectrales d'une scène et permettent une discrimination plus précise des objets que ne le permettent les images couleur (RVB) ou même les images multispectrales. À partir des images hyperspectrales, l'accès à des informations spatiales et spectrales est possible et permet de caractériser les objets de manière très fine. Bien que les potentialités de la technologie hyperspectrale apparaissent relativement grandes, l'analyse et le traitement de ces données restent des procédures complexes et difficiles. De ce fait, exploiter ces quantités importantes de données présente un défi, et la plupart des méthodes d'analyse et d'interprétation sont encore en développement.

Pour pouvoir exploiter les données hyperspectrales, les opérations de caractérisation et de classification constituent des étapes incontournables et fortement inter-dépendantes.

L'étape de caractérisation consiste à extraire des attributs pour représenter chacun des pixels de l'image. Ces attributs peuvent être issus de traitements statistiques, morphologiques ou purement radiométriques. Plusieurs travaux ont été consacrés à l'adaptation ou à l'extension des méthodes statistiques existantes d'extraction d'attributs aux cas des images hyperspectrales [1]–[4], mais la complexité de ces images nécessite des approches toujours plus élaborées pour la caractérisation des pixels, car la finesse et la pertinence de ces attributs conditionnent fortement les résultats de classification.

L'étape de classification des pixels peut être menée de manière supervisée, semi-supervisée ou non supervisée. La classification non supervisée présente plusieurs avantages par rapport à l'approche supervisée ou semi-supervisée, telle la non nécessité pour l'utilisateur, ni de préciser à l'avance les classes à discriminer ni de fournir d'échantillons d'apprentissage (des relevés de terrain). Un algorithme de classification non supervisé détecte automatiquement les classes distinctes de manière objective, ce qui réduit considérablement le risque d'erreur lié à la phase de modélisation ou à la variation temporelle des modèles, dont l'accès pour certaines applications est très difficile. Par conséquent, les méthodes supervisées et semi-supervisées ne sont pas appropriées et seront écartées dans ce travail car trop dépendantes d'informations parfois subjectives et souvent peu fiables.

Les algorithmes de classification non supervisés peuvent être classés en deux groupes : méthodes paramétriques et méthodes non paramétriques. Les méthodes paramétriques nécessitent une estimation des paramètres du mélange avant la phase de la classification. Plusieurs algorithmes ont été proposés pour l'approche paramétrique, dont le plus utilisé est certainement l'algorithme EM (Expectation-Maximisation). L'approche paramétrique présente plusieurs inconvénients en comparaison avec l'approche non paramétrique. Parmi ces inconvénients, on peut citer principalement les erreurs liées à une modélisation incorrecte des lois conditionnelles, mais aussi celles liées à l'estimation du modèle de mélange. Par conséquent, ces erreurs ne contribuent pas à l'obtention de meilleurs taux de classification.

Récemment, un algorithme de classification, appelé propagation d'affinité AP (*Affinity Propagation*), a été proposé [5]. Cet algorithme peut être utilisé en mode semi-supervisé ou non supervisé. Son principe consiste à détecter des représentants à partir des données et à former des classes autour de ces représentants de manière déterministe. L'algorithme AP a connu un vif succès en raison de ses trois principaux avantages : *i*) il peut être utilisé dans deux modes : non supervisé ou semi-supervisé, *ii*) il est insensible à l'initialisation et *iii*) il est efficace par rapport aux algorithmes semi-supervisés comme les C-moyennes [6], le FCM [7], et ISODATA [8]. Grâce à ces avantages, cet algorithme a été largement utilisé dans plusieurs domaines telles que la télédétection, la gestion des données multimédia et la reconnaissance des formes [5], [9], [10]. Toutefois, malgré son succès dans différents domaines, l'application de l'AP sur des données réelles de grande taille comme les images aériennes et satellitaires est impossible car de complexité trop élevée. De plus, la version complètement non supervisée surestime le nombre de classes [11],[12]. La recherche de solutions pour pouvoir l'appliquer sur des données de très grande taille et estimer le nombre de classes sans sur-segmentation les images s'impose.

Dans le cadre de cette thèse nous nous intéressons d'abord aux méthodes d'extraction d'attributs, puis aux méthodes non supervisées et non paramétriques pour le partitionnement des données de grande dimension (images hyperspectrales de grande taille spatiale).

Les travaux menés dans cette thèse portent sur les aspects suivants :

- L'étude des différentes méthodes de caractérisation et de classification semi-supervisée et non supervisée des pixels afin de dresser un état de l'art dans ce domaine,

- L'élaboration des solutions pour pouvoir appliquer l'algorithme de classification de propagation d'affinité aux données de grande dimension, afin d'aboutir à de meilleurs résultats de classification,
- L'application de l'approche proposée à des images naturelles de grande taille spatiale pour des applications réelles, parmi lesquelles la détection de plantes invasives.

Le mémoire est organisé en deux parties. La première partie porte sur la présentation des travaux de l'état de l'art, est composée de quatre chapitres. La seconde partie, composée de deux chapitres expose les nouveaux travaux développés dans le cadre de cette thèse ; ainsi qu'une étude expérimentale portant sur le partitionnement d'images multi et hyperspectrale de grandes tailles.

**Le premier chapitre** présente le problème posé, les objectifs à atteindre dans le cadre de cette thèse, ainsi que le contexte général dans lequel s'inscrivent les travaux menés.

**Le deuxième chapitre** dresse l'état de l'art des méthodes d'extraction d'attributs. Nous évoquons en premier lieu les attributs de la littérature, allant des moments d'ordre 1, 2 et 3 aux attributs calculés à partir des matrices de cooccurrences et des transformées en ondelettes. Nous nous intéressons plus particulièrement aux statistiques d'ordre supérieur et à leur application pour l'imagerie hyperspectrale. Nous évoquons ensuite les particularités et les atouts de ces statistiques dans le cadre de notre application (classification des pixels des images hyperspectrales) afin de justifier les voies de recherche que nous avons investies.

**Le troisième chapitre** est dédiée à l'état de l'art des méthodes de classification non paramétriques semi-supervisées et non supervisées. Cet état de l'art nous permettra d'appréhender les avantages et les inconvénients de chacune des méthodes évoquées, de proposer et de justifier les grandes lignes de l'approche de classification des pixels des images hyperspectrales présentée dans ce mémoire.

**Le quatrième chapitre** porte sur les critères d'évaluations non supervisés des méthodes de classification. Les performances des principaux critères associés à la méthode de classification retenue dans le chapitre précédent sont comparées sur des données hyperspectrales.

**Le cinquième chapitre** expose l'ensemble du travail méthodologique réalisé pendant cette thèse. Nous y présentons tout d'abord la méthodologie adoptée pour la caractérisation des pixels d'une image multi et hyperspectrale ainsi que la présentation des nouveaux attributs

de tri-occurrences calculés sur des voisinages cubiques. Ces attributs tiennent compte de l'aspect spatial et spectral des données. Une étude comparative est menée afin de tester le pouvoir discriminant de ces attributs par rapport aux autres attributs (signatures spectrales, moments statistiques, attributs d'Haralick, attributs issus des transformées en ondelettes). Plusieurs évaluations sont menées sur des images multi-composantes pour valider la démarche suivie.

Ce chapitre se poursuit par la présentation de l'approche de classification proposée pour partitionner les images de grandes tailles. Deux étapes fondamentales sont introduites pour pouvoir appliquer l'approche de classification par propagation d'affinité à des données de grande dimension : la réduction du nombre d'individus à classer, et l'optimisation et application de l'AP sur les données retenues.

Nous y détaillons les approches développées et donnons des résultats comparatifs sur des images réelles et synthétiques.

Dans **le sixième chapitre**, nous présentons les résultats de l'application de la chaîne développée sur des images réelles hyperspectrales ou multispectrales de grande dimension spatiale. Trois applications sont considérées : *i)* la l'identification de plantes invasives et non invasives, *ii)* la détection des pins ; *iii)* la classification du couvert paysager.

En **conclusion générale**, nous résumons notre contribution et tirons les conclusions sur l'approche développée. Nous proposons enfin quelques pistes d'améliorations de l'approche de partitionnement proposée.

# **Partie I**

## **État de l’art**

Dans cette partie, structurée en quatre chapitres, nous mettons tout d’abord en lumière le contexte et la problématique de l’analyse d’images et de leur partitionnement dans le premier chapitre. Nous dressons ensuite un état de l’art des deux grandes étapes nécessaires à l’interprétation du contenu informationnel des images que sont l’extraction d’attributs pertinents à partir de ces images (chapitre 2), et la classification des pixels à partir de ces caractéristiques (chapitre 3). Cette étape de classification nécessitera, dans le contexte non supervisé que nous aurons auparavant justifié et choisi, l’utilisation de mesures intrinsèques de la qualité des résultats de classification qui seront décrites et discutés dans le chapitre 4.

# Chapitre 1

## Description du contexte et problématique

Dans ce chapitre, nous évoquons en premier lieu le contexte général des travaux menés dans le cadre de cette thèse. Nous nous intéressons particulièrement à une technologie pleine de promesses : l'imagerie hyperspectrale. Après avoir évoqué les particularités et les atouts de ce nouvel outil, nous dressons un bilan des problèmes liés à son exploitation.

### 1.1. Contexte général

La présente étude concerne le partitionnement des images hyperspectrales par des méthodes de classification. L'objectif principal est de pouvoir classer les différents pixels de l'image sans aucune connaissance *a priori*.

En guise d'introduction au contexte du travail réalisé, nous présentons tout d'abord, dans cette section, l'imagerie hyperspectrale, ses caractéristiques et ses avantages. Dans un second temps, nous introduisons la problématique de la classification des pixels issus de ce type d'imagerie, et les nombreuses difficultés rencontrées dans ce domaine.

#### 1.1.1. Imagerie hyperspectrale.

Une définition formelle de l'imagerie hyperspectrale (HS) est donnée par Kruse [13]: "L'imagerie hyperspectrale consiste à acquérir des spectres pour tous les pixels d'une image, où un spectre est une mesure contiguë d'une distribution de longueur d'onde, avec une résolution suffisante pour résoudre la variabilité naturelle du système d'intérêt". Chang [14] précise cette définition en distinguant l'imagerie HS de l'imagerie couleur et multispectrale par trois caractéristiques principales (voir Figure 1).

- Les capteurs couleur ou multispectraux enregistrent une image de scène dans au minimum trois et au plus une dizaine de bandes spectrales, alors que les systèmes d'imagerie hyperspectrale acquièrent des centaines de bandes contiguës et disjointes ;
- Les systèmes multispectraux ont une résolution spectrale (longueur d'onde centrale divisée par la largeur de la bande spectrale) de l'ordre de 10, alors qu'elle est de l'ordre de 100 pour les systèmes d'imagerie HS.
- Alors que les systèmes multispectraux acquièrent les images dans des bandes de longueur d'onde larges et irrégulièrement espacées, les systèmes d'imagerie HS produisent des

bandes spectrales contiguës et régulièrement espacées, permettant d’obtenir un spectre quasi-continu pour chaque pixel.

Le Tableau 1 dresse un bilan des possibilités étendues de l’imagerie HS à travers différentes caractéristiques en comparaison avec les techniques optiques plus classiques : imagerie couleur (RVB), imagerie multispectrale (MS) [14]

Tableau 1 : Caractéristiques de l’imagerie RVB, de l’imagerie MS, et de l’imagerie HS.

Caractéristiques	Imagerie RVB	Imagerie MS	Imagerie HS
Information spatiale	$\infty$	$\infty$	$\infty$
Information spectrale	Limitée	Limitée	$\infty$
Information multi-composée	Limitée	Limitée	$\infty$
Sensibilité aux composés mineurs	Limitée	Limitée	$\infty$

L’imagerie HS permet ainsi de décrire les objets d’une manière finement détaillée grâce à de nombreux attributs.

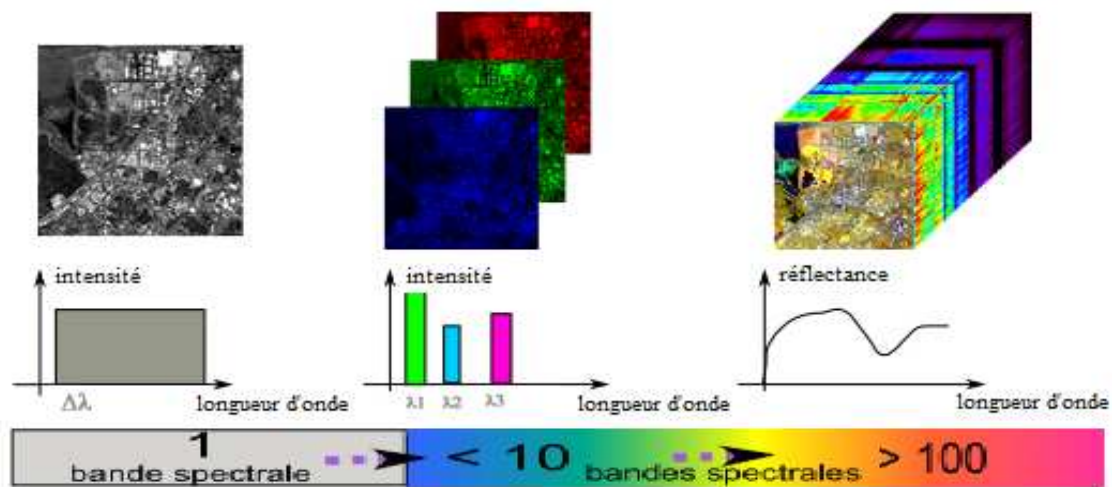


Figure 1: De la vision numérique monochrome à l’imagerie hyperspectrale [14]

### 1.1.2. Domaines applicatifs.

Les images HS fournissent donc une information plus détaillée des propriétés spectrales d’une scène et permettent d’accéder à une identification et une discrimination plus précises des objets imagés que les capteurs multispectraux.



De fait, l'émergence depuis le milieu des années 1990 de cette technologie a permis d'envisager un large spectre d'applications [15]. La miniaturisation très récente des dispositifs d'acquisition d'imagerie HS permet d'accroître encore le champ des applications potentielles [16]. Les principales applications de l'imagerie HS sont :

- la médecine (ex. : extraction du réseau veineux, détection de mélanome) [17];
- le contrôle de production (ex. : détection de défauts) [18];
- la surveillance de sites industriels (ex. : détection d'émissions polluantes) [19];
- l'agro-alimentaire (ex. : analyse de la qualité) [20];
- l'agriculture (ex. : traitement ciblé des parcelles) [21];
- la défense (ex. : détection de cibles) [22].

Les chercheurs dans ces différents domaines ont ainsi un accès à des informations spatiales et spectrales permettant de caractériser un échantillon de manière rapide et avec une résolution spatiale et spectrale très fine pour analyser une étendue d'échantillons hétérogènes complexes. Toutefois, cette multiplicité des attributs spectraux n'est pas sans poser de problèmes pour le traitement postérieur à l'acquisition.

## **1.2. Problèmes liés à la classification des données en grande dimension**

Bien que les potentialités de la technologie hyperspectrale apparaissent relativement grandes, l'exploitation de ces données est une procédure complexe et difficile. En effet, les particularités des données hyperspectrales (et leur grande dimensionnalité en particulier) ne permettent pas l'utilisation des algorithmes de classification ou de segmentation classiquement utilisés en imagerie couleur ou multispectrale. Ces particularités nécessitent en effet le développement d'algorithmes de traitement dédiés, qui permettent, d'une part, de s'affranchir du phénomène touchant les espaces de grande dimension (effet de Hughes ou malédiction de la dimensionnalité [23]), et d'autre part, de prendre en compte les dimensions spectrale et spatiale des données.

### **1.2.1. Choix des attributs - Information spatiale et spectrale**

La description du contenu des images se fait par le biais des attributs qui doivent permettre de discriminer les classes en présence. Dans le cas des données hyperspectrales, il est nécessaire d'utiliser des attributs qui soient non seulement capables de caractériser l'information spectrale, mais qui tiennent compte également de l'information spatiale.

La plupart des approches disponibles pour le traitement des données hyperspectrales ne prennent pas en compte conjointement les dimensions spatiales et spectrales des données. Les images hyperspectrales ne sont généralement pas traitées comme des matrices (images) mais comme une liste non ordonnée de spectres sans aucun arrangement spatial particulier. Ce constat a été posé par de nombreux auteurs, dont Tadjudin et Landgrebe [25]. La prise en compte de l'information spatiale et spectrale apparaît cependant très utile dans le cas de problèmes de classification très difficiles où les objets à discriminer sont spectralement très proches.

En effet, l'utilisation simultanée de l'information spectrale et spatiale est une voie pour discriminer des structures variables réalisées à partir de matériaux identiques [25]. L'utilisation d'une information de structuration des pixels telle que la taille ou la forme des régions de l'image permettra par exemple, de discriminer le toit d'une maison privée de celui d'un large immeuble. Dans le domaine de l'agriculture, la distinction entre culture et adventice est parfois très délicate en particulier pour des sols ayant déjà une couverture végétale [26]. Les auteurs expliquent alors que l'information spectrale seule n'est pas suffisante pour discriminer de façon robuste les adventices des cultures. En effet, la réflectance des plantes est extrêmement fluctuante en fonction de l'espèce, de l'état physiologique de la plante (stress hydrique, état de santé), de l'état de la feuille (orientation suivant la lumière incidente, âge, structure), de la nature et de l'état du sol, ainsi que de la structure complexe du couvert végétal. Cela nécessite l'introduction d'autres caractéristiques pour parvenir à un taux de classification correct.

### 1.2.2. Choix de la méthode de classification

La classification de données en imagerie consiste à partitionner l'ensemble des pixels de l'image en groupes homogènes et cohérents, en se basant sur les attributs de ces pixels, soit directement issus de l'acquisition (mesure de luminance, éventuellement spectrale), soit après un traitement d'extraction de nouveaux attributs liés à ces derniers et permettant l'accès à des informations de plus haut niveau structurel (attributs de texture, etc.).

Le choix de l'approche de classification est une question importante, car il influence fortement l'analyse et l'interprétation des données. De nombreuses approches de classification d'objets ont été développées dans la littérature depuis plusieurs dizaines d'années. Ces approches paramétriques ou non paramétriques peuvent être répertoriées en trois catégories : la classification supervisée, la classification semi-supervisée et la classification non supervisée.

La première catégorie exige généralement une connaissance *a priori* sur les données à classer : nombre de classes, et/ou estimation du modèle du mélange, et/ou échantillon d'apprentissage. Il s'agit ici d'apprendre à classer un nouvel individu parmi un ensemble de classes prédéfinies, déterminées à l'avance. Parmi ces méthodes, on peut citer l'analyse

discriminante linéaire de Fisher [27], les méthodes de classification basées sur l'algorithme EM (Estimation-Maximization) [28], les méthodes neuronales [29], les méthodes à noyau de type SVM (Support Vector Machines) [30]. Ces approches ont montré plusieurs limites pour le traitement des données en grandes dimensions. Par exemple, les méthodes s'appuyant sur des modèles imposés de distribution conditionnelle des classes, comme des distributions multinormales, sont souvent largement insuffisantes pour représenter fidèlement la réalité. Par ailleurs, les méthodes utilisant un jeu d'apprentissage ne peuvent conduire à de bons résultats que si le jeu d'apprentissage est absolument fiable, ce qui est loin d'être le cas dans beaucoup d'applications. Les méthodes de classification semi-supervisées exigent quant à elles un minimum de connaissances *a priori*, comme le nombre de classes à utiliser pour le partitionnement des objets.

Remédiant aux limites de ces approches, la classification non supervisée et non paramétrique permet un traitement de tout type de données sans aucune hypothèse faite à l'avance, ni supervision. À partir d'une population d'objets, il s'agit d'extraire des classes ou groupes d'individus présentant des caractéristiques communes, le nombre et la définition des classes n'étant, cette fois, pas donnés *a priori*. En général, l'approche non supervisée présente plusieurs avantages par rapport à l'approche supervisée [31] :

- elle permet la détection automatique des classes distinctes, qu'elles soient connues à l'avance ou non, ce qui réduit considérablement le travail de l'utilisateur.
- les caractéristiques des objets à classer pouvant varier dans le temps, l'approche non supervisée est un excellent moyen pour suivre ces changements.

Le choix de la méthode de classification est donc crucial dans le but d'obtenir de meilleurs résultats de classification. Pour cette raison, une étude approfondie sur les méthodes de classification non paramétriques et non supervisées sera présentée dans la section suivante. Cette étude nous permettra de choisir la méthode qui sera développée par la suite.

### 1.3. Conclusion

Dans ce chapitre introductif, nous avons évoqué les propriétés et les atouts des images hyperspectrales dans une problématique de classification de scènes complexes. L'avenir de ce type d'images apparaît prometteur mais nécessite un investissement de recherche pour résoudre des problèmes liés à l'exploitation de ces données. Dans le cadre de cette thèse, nous nous intéresserons plus particulièrement aux grandes étapes de la chaîne d'analyse et d'interprétation des images que sont la caractérisation et à la classification des pixels de ce type de données.

Dans ce chapitre, deux problématiques afférentes à l'exploitation des données hyperspectrales ont été évoquées, à savoir :

- (1) Comment choisir les attributs les plus pertinents en vue de la classification,
- (2) Comment choisir l'approche de classification la plus appropriée pour mieux partitionner les données de grandes dimensions.

Le chapitre suivant présente un état de l'art sur les attributs les plus pertinents pour la description des images. Il est suivi d'un chapitre dédié à un état de l'art sur les méthodes de classification. Nous nous intéressons uniquement aux approches de classification non supervisées et semi-supervisées non paramétriques. Cet état de l'art nous permettra d'appréhender les avantages et les inconvénients de chacune des méthodes évoquées, de justifier les grandes lignes de l'approche de classification des pixels des images HS proposée dans ce mémoire. Dans le cadre non supervisé que nous avons choisi, nous détaillons dans le dernier chapitre de cette partie les outils de mesure de la qualité intrinsèque d'une classification.

## Chapitre 2

### Caractérisation des individus

#### 2.1. Introduction

Pour pouvoir analyser et interpréter le contenu informationnel des données, l'extraction d'attributs et la classification sont deux étapes indispensables. La description du contenu des images se fait par le biais des attributs qui doivent permettre de discriminer les classes en présence.

L'étape d'extraction d'attributs regroupe un ensemble de techniques de calcul de paramètres pour caractériser les pixels de l'image à partitionner. Il est bien connu que la pertinence de ces attributs conditionne fortement les résultats de classification.

Avant de passer en revue dans le Chapitre 3 les principales méthodes de classification non paramétriques, non supervisées et semi-supervisées, nous présentons ici un état de l'art sur les principaux attributs utilisés pour la caractérisation des pixels (individus).

Il existe de nombreuses techniques de caractérisation d'un pixel dans son environnement spatial. Par exemple, les attributs peuvent être issus de l'histogramme de l'image, de l'image gradient, de la matrice de cooccurrences, de la matrice des longueurs de plage, des coefficients des filtres de Gabor ou des coefficients de la transformée en ondelettes [32]. Les méthodes statistiques d'extraction d'attributs étudient la relation entre un pixel et ses voisins et définissent des paramètres discriminants de la texture en se basant sur des outils statistiques. Généralement, elles sont utilisées pour caractériser des structures fines, sans régularité apparente.

Dans le cas des images multicomposantes (multi ou hyperspectrales), la richesse d'information fait qu'il est possible de définir trois grandes familles d'attributs :

- **spatiaux** : calculés sur une fenêtre d'analyse suivant les dimensions spatiales,
- **spectraux** : calculés sur une fenêtre d'analyse suivant la dimension spectrale,
- **spatiaux et spectraux** : calculés sur une fenêtre d'analyse cubique suivant les dimensions spatiales et spectrale.

Dans le cas des données hyperspectrales, il est nécessaire d'utiliser des attributs qui soient non seulement capables de caractériser l'aspect spectral, mais qui tiennent compte également de l'information spatiale. La prise en compte de l'information spatiale apparaît très utile dans le cas de problèmes de classification complexes où les objets à discriminer sont spectralement très proches. En effet, l'utilisation simultanée de l'information spectrale et spatiale est une voie pour discriminer des structures variables réalisées à partir de matériaux identiques [26].

Rappelons que la démarche globale adoptée dans ce travail est d'éviter l'utilisation d'information *a priori* dans l'élaboration de méthodes de classification. Aussi, nous ne présentons ici que les attributs les plus pertinents, ne nécessitant aucune hypothèse d'utilisation.

Le choix d'attributs est une étape cruciale de laquelle dépendent fortement les performances du processus de classification. Ce chapitre est donc dédié aux méthodes d'extraction d'attributs. Après avoir rappelé les caractéristiques des moments d'ordre 1, 2 et 3, des attributs calculés à partir des matrices de cooccurrences et des longueurs de plages, et de la transformée en ondelettes, nous nous intéressons plus particulièrement aux statistiques d'ordre supérieur et à leur application pour l'imagerie hyperspectrale (HS). Nous évoquons ensuite les particularités et les atouts de ces statistiques dans le cadre de notre application (classification des pixels des images HS) afin de justifier les voies de recherche que nous avons investies.

## 2.2. Attributs spatiaux

Pour la caractérisation des pixels des images monocomposantes, les attributs les plus utilisés sont les moments statistiques [34], les attributs de cooccurrences (attributs d'Haralick) [35] [36], les coefficients d'ondelettes [37] et les statistiques d'ordre supérieur [38].

### 2.2.1. Les moments statistiques

Les moments statistiques sont calculés soit à partir de l'histogramme d'une image entière, soit en considérant l'ensemble des pixels dans une fenêtre donnée.

Soit  $a_j$  le vecteur d'attributs caractérisant un pixel  $x_j$ . Dans le cas d'une image monocomposante, on pose  $a_j = l_j$  le niveau de gris du pixel  $x_j$ .

Soit  $N_v$  le nombre de pixels d'une fenêtre d'analyse  $W$  avec  $x_j \in W$ .

Les quatre premiers moments calculés localement pour caractériser un pixel  $x_i$  en tenant comptes des pixels voisins d'une fenêtre  $W$  centrée sur le pixel  $x_i$  sont : moyenne, variance, skewness (ou coefficient de dissymétrie) et kurtosis (ou coefficient d'aplatissement) [34]. Ils s'écrivent respectivement :

$$\mu(x_i) = \frac{1}{N_v} \sum_{j=1}^{N_v} \ell_j \quad (2.1)$$

$$Var(x_i) = \frac{1}{N_v} \sum_{j=1}^{N_v} (\ell_j - \mu(x_i))^2 \quad (2.2)$$

$$Skewness(x_i) = \frac{1}{N_v} \sum_{j=1}^{N_v} (\ell_j - \mu(x_i))^3 \quad (2.3)$$

$$Kurtosis(x_i) = \frac{1}{N_v} \sum_{j=1}^{N_v} (\ell_j - \mu(x_i))^4 \quad (2.4)$$

Ces attributs, couramment utilisés pour la caractérisation des pixels, ont été aussi utilisés pour estimer la dynamique ou la rugosité des textures [39], [40]. Ils permettent une classification fine s'ils sont utilisés sur des fenêtres de petite taille (3x3) et intègrent bien les détails de la texture sur des fenêtres de taille encore plus grande (7x7, 9x9, 11x11). Sur des fenêtres de grande taille, ces attributs permettent une bonne discrimination des classes, mais ils introduisent des effets de bord propres pouvant conduire à des taux de classification erronés.

Une étude menée par Rosenberger et Chehdi [41] pour l'analyse des images texturées a montré que les moments d'ordre 1 et 2 seuls ont permis d'obtenir 72.3% de taux moyen de classification correcte (noté ACCR : Average Correct Classification Rate) sur 50 échantillons de 21 textures différentes, représentatives de l'album de Brodatz [42], la classification étant opérée par le critère de minimisation de la distance euclidienne. Il est également montré que l'utilisation conjointe des quatre moments permet d'atteindre un taux moyen de bonne classification de 92.6%. De plus, leur utilisation conjointe avec d'autres attributs (paramètres

d'un modèle de textures ou attributs de cooccurrences), permet d'augmenter le taux de bonne classification de 5.4%.

### 2.2.2. Les attributs de la matrice de cooccurrences

Dans le but de décrire une région texturée, un ensemble d'attributs basé sur le calcul des matrices de cooccurrence des niveaux de gris a été proposé par Haralick [43]. Les matrices de cooccurrence constituent l'outil le plus populaire pour l'identification et la classification de textures [43]–[45]. Ils ont été utilisés de deux manières, soit directement sur l'image en procédant par une étape de réduction des niveaux de gris (à 32 ou 64 niveaux) ou bien sur les coefficients d'ondelettes [46], [47].

La matrice de cooccurrence consiste à estimer les fréquences d'apparition de couples de niveaux de gris de pixels dans une zone de l'image. Les éléments de cette matrice notés  $M_{d,\theta}(\ell_i, \ell_j)$  indiquent la fréquence d'occurrence d'un couple de niveaux de gris  $(\ell_i, \ell_j)$  séparés par une distance  $d$  et situés sur une droite faisant un angle  $\theta$  par rapport à l'axe horizontal avec  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ .

Dans le cas d'une image monocomposante, les statistiques d'ordre 2 ou les attributs d'Haralick mesurent la distribution des couples de niveaux de gris  $(\ell_i, \ell_j)$  conjointement présents dans une image en fonction de l'éloignement des pixels. Cette méthode donne généralement lieu au calcul d'une matrice  $M$  pour une direction définie par un angle  $\theta$  et pour une distance  $d$  séparant les sites du couple. Pour une image  $I$  ayant  $N_\ell$  niveaux de gris distincts, la matrice de cooccurrences admet  $N_\ell^2$  éléments pour chaque direction  $\theta$ . Même si on considère que cette matrice est symétrique (le nombre de transitions du niveau de gris  $\ell_i$  vers le niveau  $\ell_j$  est égal au nombre de transitions du niveau  $\ell_j$  vers le niveau  $\ell_i$ ), le nombre d'éléments utiles de la matrice  $(N_\ell \times (N_\ell + 1)/2)$  est généralement beaucoup trop important pour que la matrice soit directement utilisée pour caractériser les pixels de l'image. Une première réduction de ce nombre consiste par exemple à réduire le nombre de niveaux de gris de l'image par une fonction linéaire. Haralick et al. [35] ont proposé un ensemble de quinze attributs (cf. Annexe A) calculés à partir des matrices de cooccurrences pour caractériser les textures de façon efficace et surtout concise.



L'étude menée dans [41] a permis également l'analyse de la redondance entre plusieurs types d'attributs : cooccurrences, extrema locaux, longueur de plages, histogrammes locaux, intégrale curviligne, modèle d'autocorrélation et moments d'ordre  $k$ . Sur la même image constituée de textures de l'album de Brodatz [42], cette étude a montré que neuf attributs d'Haralick sur quinze ne sont pas redondants suivant le critère de corrélation. Le taux moyen de bonne classification est similaire en utilisant les neuf ou les quinze attributs (86.6%). Ceci permet donc de réduire l'espace de mesures, car l'utilisation conjointe d'attributs redondants n'apporte aucune information complémentaire et pertinente pour la discrimination des textures entre elles.

Cette étude a montré que l'association des attributs de cooccurrences, les moments d'ordre 1 à 4, donne un taux de bonne classification autour 98,8% en appliquant la méthode de classification supervisée utilisant le critère de minimisation de la distance euclidienne.

### 2.2.3. Les attributs des longueurs de plages

Galloway [33] a introduit la méthode de la matrice de longueur de plages qui consiste à rechercher les plages (ou segments de pixels) de même niveau de gris.

Une plage est définie comme un ensemble consécutif de points ayant des niveaux de gris similaires  $\ell$ , dans une direction donnée  $\theta$ .

Pour chaque orientation, une matrice  $PL(\ell, L, \theta)$  de longueurs de plages de taille  $N_l \times L$  est calculée, avec  $N_l$  le nombre de niveaux de gris de l'image ou de la fenêtre à considérer, et  $L$  la longueur maximale d'une plage.

Les cinq attributs de la matrice de longueur de plages utilisés sont précisés en annexe A.

### 2.2.4. Discussion

Dans cette section nous avons présenté un état de l'art des principaux attributs spatiaux les plus pertinents, ne nécessitant aucune hypothèse d'utilisation. Chaque type d'attributs a ses caractéristiques propres et peut s'appliquer à n'importe quel type d'images monochromes. Vu l'importance du nombre de paramètres qui peuvent être utilisés (voir Annexe A), une étude a été menée dans [39], [41] pour analyser la redondance ou la corrélation des attributs adaptés à la caractérisation de la texture les plus utilisés. Ces travaux ont permis

de réduire l'espace de mesures car l'utilisation conjointe d'attributs redondants n'apporte aucune information complémentaire.

L'étude a permis de dégager les conclusions suivantes:

- les attributs de la matrice de cooccurrence sont relativement redondants. En effet, seuls 9 de ces 15 attributs sont non redondants,
- les attributs de la matrice de longueur de plages sont un peu plus redondants que ceux de la matrice de cooccurrence et sont moins nombreux. Seuls 3 attributs sur 5 sont moins redondants,
- les moments d'ordre 1 à 4 sont complémentaires aux attributs définis et donnent de bons résultats.

Il convient de noter que :

- les attributs extraits des matrices de cooccurrences permettent une caractérisation correcte des pixels dans le cas des images texturées simples mais restent insuffisants pour représenter finement des textures complexes.
- les statistiques d'ordre supérieur montrent l'apport important pour la caractérisation des pixels des images (meilleur que les statistiques d'ordre deux), ce qui contribue à une meilleure discrimination. Ces attributs sont complémentaires à ceux extraits des matrices de cooccurrences.
- Les attributs présentés ont été utilisés principalement pour les images monochromes. Pour le cas des images hyperspectrales, ils peuvent être calculés bande par bande.

Nous présentons dans la section 2.7 les différentes extensions de ces attributs pour les images multispectrales et hyperspectrales.

### 2.3. Attributs issus de la transformée en ondelettes

Les transformations orthogonales permettent de représenter et d'étudier les signaux et les images dans un autre espace que celui d'origine. Les attributs issus d'une transformation orthogonale sont calculés généralement par transformation locale de l'image et peuvent être utilisés pour le partitionnement des images présentant des textures. Certaines des transformations (transformées en sinus, cosinus, Hadamard, Fourier, et ondelettes) sont à

valeur dans des espaces qui peuvent être propices à l'extraction d'attributs pertinents. C'est le cas notamment de la transformée en ondelettes qui permet par exemple une meilleure description des signaux non stationnaires [37].

#### *Transformée en Ondelettes :*

L'analyse multi-résolution consiste à décomposer l'image sur une gamme très étendue d'échelles, opération que nous pouvons comparer à une cartographie (Figure 2). A chaque échelle, l'image est remplacée par l'approximation la plus adéquate que l'on puisse y tracer. En allant des échelles les plus grossières vers les plus fines, nous accédons à des représentations de plus en plus précises de l'image donnée jusqu'à l'échelle maximale correspondant à l'image originale. L'analyse s'effectue en calculant ce qui diffère d'une échelle à l'autre, c'est à dire les détails à une résolution donnée.

En corrigeant une approximation encore assez grossière, ces détails permettent d'accéder à une représentation d'une qualité meilleure. Parmi les méthodes multi-résolutions les plus utilisées dans la littérature, nous allons nous intéresser à la décomposition en ondelettes [48].

La transformée en ondelettes a été introduite par Jean Morlet en 1983 pour faciliter l'étude des signaux sismiques [32]. Cette transformée permet la représentation de l'évolution temporelle d'un signal monodimensionnel. Adaptée au traitement de données bidimensionnelles [48], la transformée en ondelettes donne accès à une représentation spectrale multirésolution de l'image initiale. Cette représentation est utilisée en analyse de textures dans [37] [49][50]. La transformée en ondelettes entre dans le cadre des méthodes multi-échelle.

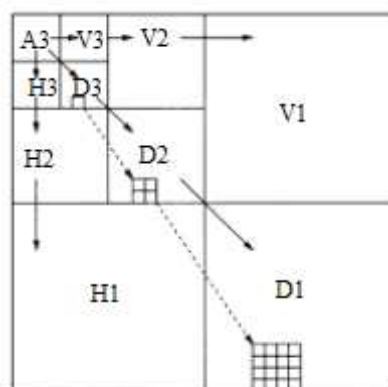


Figure 2 : Exemple de décomposition en 3 niveaux d'ondelettes

Nous introduisons ici quelques concepts de base relatifs à la transformation en ondelettes. On peut trouver plus de détails sur la théorie des ondelettes dans plusieurs travaux,

notamment ceux de Meyer [51], de Mallat [48], et de Truchetet [52]. Les ondelettes sont des fonctions dont la forme générale est donnée par la formule suivante :

$$\Psi_{h,\tau}(t) = \frac{1}{\sqrt{h}} \Psi\left(\frac{t-\tau}{h}\right), \text{ avec } (h,\tau) \in \mathbb{R}^+ \times \mathbb{R} \quad (2.5)$$

où le paramètre  $h$  est le facteur d'échelle,  $\tau$  le paramètre de translation, et  $\Psi$  est une fonction appelée ondelette mère.

L'analyse par ondelettes est un outil mathématique capable de transformer un signal d'énergie finie dans le domaine spatial en un autre signal d'énergie finie dans le domaine spatio-fréquentiel. Les composantes de ce nouveau signal sont appelées les coefficients d'ondelettes. Ces coefficients renseignent sur la variation locale des niveaux de gris autour d'un pixel donné de l'image. Ils sont d'autant plus grands que cette variation est importante. Contrairement à la transformée de Fourier, l'analyse par ondelettes est localisée dans l'espace et en fréquence, ce qui la rend très utile pour l'analyse des signaux non stationnaires. En 1989, Mallat [48] a proposé un schéma d'analyse multi-résolution basé sur la transformation en ondelettes. L'algorithme décompose une image en un ensemble de trois images de détails et une image d'approximation. À chaque niveau de décomposition, la taille de l'image et des images de détails associés d'approximation est réduite d'un facteur deux par rapport au niveau antérieur.

Deux paramètres doivent être fixés pour la transformée en ondelettes :

- **Le choix de l'ondelette mère:** dans plusieurs études [50], [53]–[56], les auteurs ont essayé de donner une réponse générale au problème du choix de l'ondelette mère, mais il n'y a pas de solution globale et le choix de l'ondelette dépend essentiellement du type d'application visé (compression, classification, etc.). Pour ce qui est de la classification, les ondelettes de Daubechies [54]–[56] sont souvent utilisées pour l'analyse de la texture car elles permettent une bonne localisation dans les domaines spatial et fréquentiel et sont sensibles à la singularité locale et à la corrélation de l'image.
- **Le choix de la transformation :** du fait que l'algorithme décompose une image en un ensemble d'images de détails et une image d'approximation, à chaque niveau de décomposition la taille de l'images d'approximation est réduite d'un facteur deux. Pour des besoins particuliers en traitement d'image, cette réduction de taille peut constituer un handicap, car la correspondance pixel à pixel entre les différentes

résolutions n'est pas préservée (Figure 3). Afin de pallier ce problème, la transformation bicubique [55] est souvent utilisée.

Les attributs les plus utilisés pour caractériser la texture dans le plan spatio-fréquentiel sont l'énergie et la moyenne [57]. L'attribut énergie est une mesure locale de la distribution des coefficients d'ondelettes en fonction de la fréquence de l'orientation et de l'échelle. Elle a été utilisée avec succès pour des problèmes de classification et de segmentation des textures [54], [58]. Ces deux attributs de texture sont calculés comme suit :

Soit  $c_i$  la valeur de la transformée du pixel  $x_i$  pour les sous bandes d'approximations sur les différents niveaux de décompositions.

Les attributs moyenne et énergie calculés sur une fenêtre  $W$  de taille  $N_v$  centrée sur la valeur de la transformée du pixel  $x_i$  sont donnés respectivement par :

$$\mu_w = \frac{1}{N_v} \sum_{i=1}^{N_v} c_i \quad (2.6)$$

$$Ene_w = \frac{1}{N_v} \sum_{i=1}^{N_v} c_i^2 \quad (2.7)$$

Les moments statistiques ainsi que les attributs de cooccurrences à partir des coefficients d'ondelettes ont été utilisés avec succès pour des problèmes de classification et de segmentation des textures de l'album de Brodatz [46], [47]. La procédure de classification est obtenue par une simple comparaison des attributs calculés par bloc. La localisation précise des frontières entre régions est réalisée par l'application d'un algorithme de squelettisation.

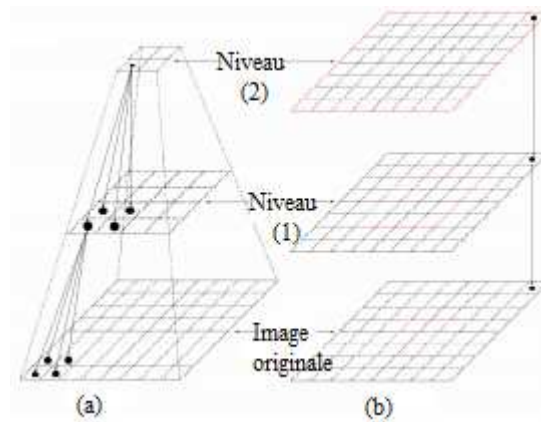


Figure 3 : Évolution du voisinage en fonction de la résolution

## 2.4. Les statistiques d'ordre supérieur

Les statistiques d'ordre supérieur, autrement dit les moments et cumulants d'ordre supérieur à deux, sont essentiellement utilisées pour combler les insuffisances de caractérisation des individus par des statistiques d'ordre deux. Elles donnent une description plus complète des données et de leurs propriétés et contribuent ainsi à une meilleure discrimination [74-75].

L'un des fondateurs de l'analyse et de la synthèse d'image, André Gagalowicz [59] [60], a montré que les statistiques d'ordre supérieur permettent une meilleure description des images que les statistiques du second ordre. En effet, il a montré que deux images présentant des microstructures différentes, ont des statistiques d'ordre deux identiques mais des statistiques d'ordre supérieur différentes. D'autres auteurs [61], [62] ont confirmé cette conclusion pour la caractérisation des pixels des images texturées synthétiques et réelles. Cela implique que pour une meilleure description d'une texture, les statistiques d'ordre supérieur doivent être introduites.

Cruz et al. [38] ont montré également que les bi-corrélations et tri-corrélations, donnent des caractéristiques plus riches (plus discriminantes) que les moments d'ordre deux pour la représentation des microtextures, car elles estiment des corrélations entre 3 ou 4 pixels dans un voisinage donné autour du pixel courant. Ils ont montré également que la bi-corrélation est moins pertinente que la tri-corrélation pour la caractérisation des textures analysées. Coroyer [62] a montré sur un échantillon de la base d'images de Brodatz [42] que les statistiques d'ordre trois décrivent mieux les micro-textures dont les distributions sont plutôt non-gaussiennes. Sur 504 images de taille 64x64 pixels, les statistiques d'ordre supérieur donnent un taux de bonne classification de 98.4% contre 95.4% pour les cooccurrences en utilisant une approche de classification neuronale. Ce résultat a été également confirmé sur d'autres bases d'images de taille 32x32 et 16x16 pixels par le même auteur.

## 2.5. Attributs spectraux

En plus des attributs liés directement à la signature spectrale de chaque pixel, d'autres attributs spécifiques à l'imagerie multicomposantes ont été proposés comme la dérivée spectrale [63] qui peut être estimée par l'obtention de l'information sur la pente de la courbe de réflectance. Les données hyperspectrales acquises dans des conditions réelles sont souvent

bruitées. Le bruit peut potentiellement être amplifié lorsque les dérivées sont calculées sur ces données. Par conséquent, il devient nécessaire de filtrer les données en réflectance par un filtre qui supprime de telles perturbations.

La première dérivée spectrale est calculée comme suit :

$$\frac{\partial R}{\partial \lambda} \approx \frac{R(\lambda_j) - R(\lambda_i)}{\Delta \lambda} \quad (2.8)$$

avec  $\Delta \lambda = \lambda_j - \lambda_i$  la différence des longueurs d'ondes  $\lambda_i$  et  $\lambda_j$  correspondant aux bandes  $i$  et  $j$ , avec  $\lambda_i < \lambda_j$ .

La dérivée seconde s'écrit :

$$\frac{\partial^2 R}{\partial \lambda^2} \approx \frac{R(\lambda_k) - 2R(\lambda_j) + R(\lambda_i)}{(\Delta \lambda)^2} \quad (2.9)$$

avec  $\Delta \lambda = \lambda_k - \lambda_j - \lambda_i$ ;  $\lambda_k > \lambda_j > \lambda_i$

L'intégration de l'information liée à la dérivée spectrale a permis d'obtenir une meilleure séparation entre les classes que l'utilisation directe des signatures spectrales [63]. Toutefois, l'utilisation de la signature spectrale directe [64], [65] pour classer les pixels donne des résultats très satisfaisants pour de nombreuses applications.

## 2.6. Attributs spatiaux et spectraux

Plusieurs travaux ont montré que l'extraction des caractéristiques reste un problème difficile pour les images couleur [66], multi ou hyperspectrales, où les attributs des pixels de l'image prennent leurs valeurs dans un espace multidimensionnel. Pour ce faire, beaucoup de travaux ont été proposés pour l'analyse des attributs dans le cas des images multispectrales et hyperspectrales [67]–[71]. Ces travaux ont été menés sur l'utilisation de filtres de Gabor [72], les matrices de cooccurrences [2], [3], [73]–[76] et la morphologie mathématique [67], [77] – [79]. L'objectif de cette étude n'est pas de réaliser une présentation exhaustive des méthodes afférentes à chacune des catégories évoquées ci-dessus mais de présenter les approches récentes qui ont suscité le plus grand nombre de travaux.

Une première généralisation des matrices de cooccurrences pour les images couleur [73] a été proposée en 1996, le calcul des attributs d'Haralick (énergie, entropie, contraste, corrélation, homogénéité) se faisant sur les composantes couleur des pixels. Par analogie au calcul de la fréquence d'apparition d'un couple de valeurs de niveaux de gris, ici le spectre

couleur associé au couple de pixels est utilisé. L'application de l'approche proposée sur des images multispectrales synthétiques à des fins de classification a montré son pouvoir discriminant (83.81% contre 60% pour les attributs calculés bande par bande) dans le cas de textures qui ont la même couleur et des structures différentes ou d'une même structure présentant des couleurs différentes. Le classifieur utilisé est celui des  $k$  plus proches voisins (KNN) qui est un classifieur supervisé.

Dans [80], Ghoneim et al. présentent une extension du calcul des matrices de cooccurrences pour la discrimination des tumeurs sur des données IRM (données 3D). L'extension se présente dans la façon dont sont disposés les deux voxels utilisés comme le montre la Figure 4. Les deux valeurs utilisées pour le calcul de la densité de probabilité conditionnelle se situent sur deux plans différents. Les résultats de discrimination des tumeurs par la méthode LDA (Linear Discriminant Analysis) [27] utilisant les attributs d'Haralick (énergie, contraste, corrélation, entropie, moment des différences inverses, moment diagonale) montrent que les cooccurrences calculées dans un cube permettent de mieux discriminer les tumeurs que les cooccurrences calculées seulement suivant les directions spatiales. Certaines tumeurs sont reconnues à 100% contre 75% pour les cooccurrences classiques.

Notons que cette approche est appliquée sur des données d'imagerie 3D, toutefois elle peut être facilement étendue à des données hyperspectrales.

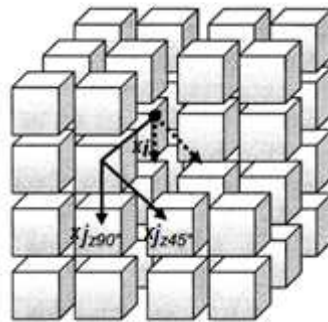


Figure 4 : Représentation de deux voxels  $x_i$  et  $x_j$  pour une distance  $d=1$  et angle  $\theta=45^\circ$ ,  $\theta=90^\circ$

Plus récemment, une autre piste de calcul d'attributs de cooccurrences pour la classification des pixels des images couleur a été étudiée [2]. Les six attributs d'Haralick (homogénéité, contraste, corrélation, entropie, homogénéité locale) sont calculés à la fois entre et au sein des trois bandes spectrales (RVB) après réduction par quantification des niveaux de gris par bande spectrale. Six matrices de cooccurrences sont nécessaires pour le calcul des 30 attributs. Ces attributs sont ensuite utilisés conjointement avec les attributs de couleurs (moyenne et l'écart



type des 3 bandes) pour la classification de 54 images couleurs de la base de données de Vistex et Outex [2]. Les résultats ont montré que les meilleurs pourcentages de bonne classification sont obtenus avec l'ensemble des attributs (méthode multispectrale et méthode spatiale classique) avec des taux moyens de bonne classification respectivement de 97,9% et 94,9%. Le KNN est utilisé comme classifieur. Pour chaque image, la moitié des pixels a été utilisée pour l'étape d'apprentissage, et l'autre moitié a servi en tant que données de test. Cette étude montre que les attributs proposés sont complémentaires à ceux de la cooccurrence classique calculée sur chaque bande (Figure 5), alors que leur combinaison augmente le taux de classification pour différentes images testées.

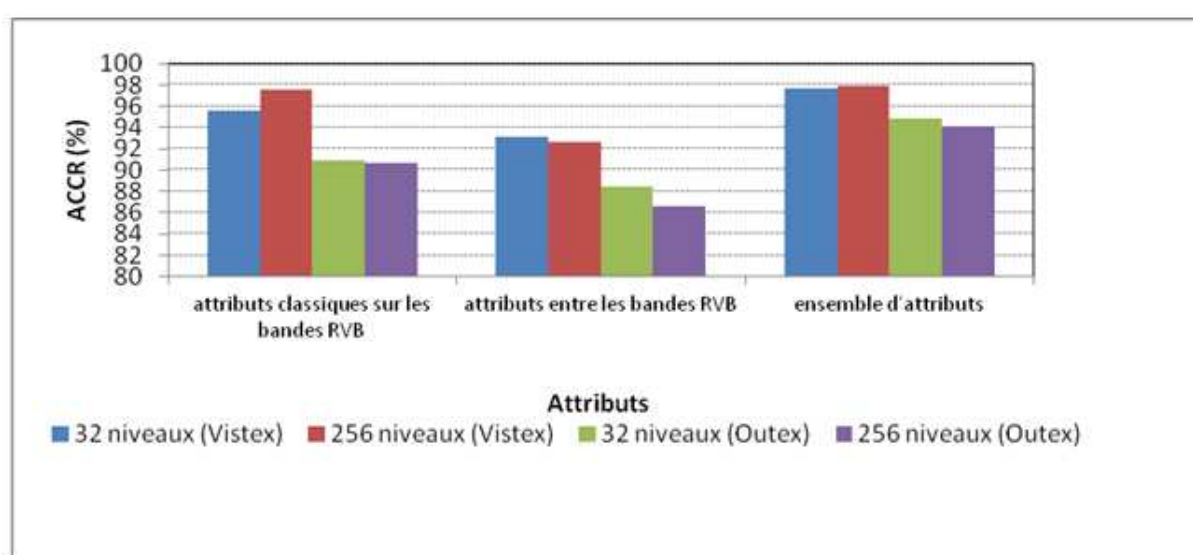


Figure 5: Taux de bonne classification en fonction des différents attributs.

Pour calculer les cooccurrences dans un volume, une étude intégrant le voisinage d'un voxel a été introduite dans [3]. Le voisinage est choisi par l'algorithme d'analyse de la semi variance 3D (les tailles retenues sont 3x3x3 et 3x3x5). Comme les méthodes déjà citées, cette approche est basée sur le calcul de la densité de probabilité du second ordre. L'application des attributs (contraste, entropie, homogénéité, second moment angulaire) à l'entrée d'un classifieur supervisé SVM pour le partitionnement de deux images hyperspectrales, Hyperion (481x255x45) et ISIS (800x800x190), a montré de meilleurs résultats que les attributs de cooccurrences classiques : 87% contre 76% pour Hyperion et 78% contre 70% pour ISIS.

Dans [74], [75], [81] Khelifi et al. ont développé une stratégie de calcul des matrices de cooccurrences aux cas des images multicomposantes qui tient compte à la fois de l'information spectrale et spatiale nommé SSGLDM (*spatial and spectral gray level*

*dependence method*). Cette méthode se base sur l'estimation des densités de probabilité du second ordre pour des vecteurs spectraux (Figure 6) suivant une distance et une direction données ( $d$  et  $\theta$ ) ; c'est une généralisation du calcul des matrices de cooccurrences classiques (GLCM) aux cas des images multi-composantes. La méthode de calcul proposée reste une méthode du second ordre puisqu'elle cherche la fréquence d'apparition d'un couple de vecteurs. Sept attributs d'Haralick ont été extraits pour caractériser l'image entière afin de la classer suivant le degré de malignité de la tumeur observée : second moment angulaire, entropie, inertie, homogénéité locale, asymétrie, proéminence, corrélation. Ces attributs ont tous été calculés pour une distance  $d=1$  et quatre directions  $\theta=0^\circ$ ,  $\theta=45^\circ$ ,  $\theta=90^\circ$  et  $\theta=135^\circ$ . Elle a été appliquée avec succès à des images multispectrales (16 bandes spectrales : [500, 650] nm avec un pas de 5 nm) texturées de cancer de la prostate en utilisant la méthode de classification SVM (96% contre 91.8% pour les cooccurrences calculés bande par bande). Une étude comparative a été menée par les auteurs pour évaluer la pertinence de ces paramètres en réduisant le nombre de bandes spectrales et les niveaux de gris. Comme le montre la Figure 7, la réduction du nombre des niveaux de gris à 32 niveaux au lieu de 64 donne dans la majorité des cas (4 cas sur 6) de meilleurs résultats de classification. Le meilleur taux est obtenu pour un nombre de bandes égal à trois et pour 32 niveaux de gris (97.76%). De plus, l'application de la méthode SSGLDM sur un sous-ensemble de bandes donne sur l'ensemble des tests de meilleurs résultats que ces mêmes attributs calculés à partir de matrices de cooccurrences classiques (GLCM). Toutefois, le taux décroît (moins bon que la GLCM) si on l'applique sur l'ensemble des bandes (94.87% pour la GLCM sur une seule bande contre 93.91% pour la SSGLDM sur 16 bandes et pour une réduction à 32 niveaux).

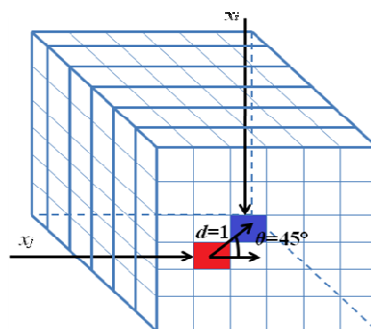


Figure 6: Représentation spectrale de deux pixels  $x_i$  et  $x_j$  pour une distance  $d=1$  et  $\theta=45^\circ$

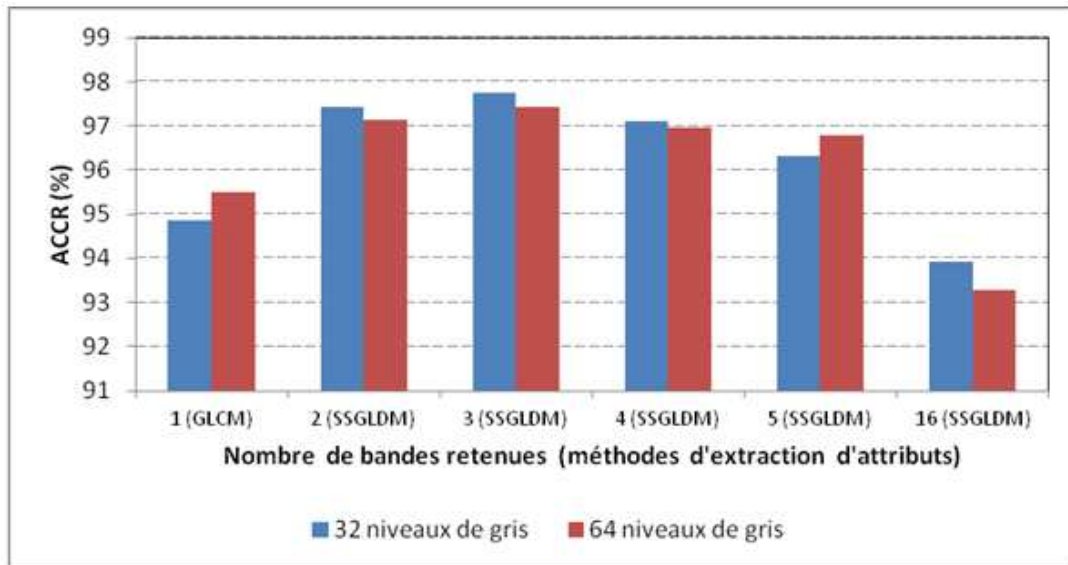


Figure 7: Taux de bonne classification en fonction des différents attributs et du nombre de bandes retenues.

## 2.7. Conclusion

L'analyse des travaux réalisés dans la littérature sur la caractérisation des pixels des images mono ou multicomposantes (couleur, multi-hyperspectrales) conduit aux conclusions suivantes :

- Les statistiques d'ordre supérieur sont les plus adaptées pour la caractérisation des scènes complexes et ont montré leur pouvoir discriminant par rapport aux attributs de texture calculés à partir des matrices de cooccurrences. Dans certains cas, la combinaison de ces deux types d'attributs améliore le taux de discrimination [41].
- La transformée en ondelettes peut également constituer un outil pertinent pour la représentation de la texture à diverses échelles [46], [47], [56], et optimiser le partitionnement des images en l'associant aux attributs statistiques.

Actuellement, le problème de la caractérisation des pixels des images hyperspectrales n'est pas complètement résolu car les deux informations spatiale et spectrale ne sont pas finement et conjointement exploitées. De plus, l'utilisation des statistiques d'ordre supérieur avec l'exploitation de l'information spatiale et spectrale des données hyperspectrales présente une voie de recherche prometteuse pour la caractérisation des pixels des images hyperspectrales.

Dans ce cadre, nous proposons dans la seconde partie de cette thèse une extension du calcul des matrices de tri-occurrences pour la caractérisation des données hyperspectrales.

## Chapitre 3

# Classification semi-supervisée et non supervisée

### 3.1. Introduction

Le chapitre précédent portant sur l'extraction des attributs a permis d'évoquer l'intérêt de combiner les aspects spatiaux et spectraux dans les approches de caractérisation des données hyperspectrales pour mieux partitionner le contenu d'une image. Dans le présent chapitre nous allons nous intéresser à l'étape de classification permettant de partitionner les individus ou objets à partir des attributs qui les représentent. Pour cette étape incontournable dans l'exploitation et l'interprétation des données, il existe trois grandes familles de méthodes : les méthodes supervisées, semi-supervisées et non supervisées.

La classification supervisée consiste à inférer, à partir d'un échantillon de données classées ou d'apprentissage, une procédure de classification. Plus clairement, il s'agit de modéliser une relation entre des informations déjà classées et une information cible qu'on veut classer. On suppose qu'il existe déjà une classification, donc un nombre connu de classes, où chaque groupe d'instance (classe) possède déjà un profil particulier. Les méthodes supervisées nécessitent donc de l'utilisateur une ou plusieurs interventions.

Si par ailleurs, seul le nombre de classes est connu au départ, on parle de classification semi-supervisée.

Si, maintenant on désire réaliser une classification sans connaissance *a priori* du nombre de classes ou de leurs caractéristiques, on effectue alors une classification dite non supervisée. Contrairement aux méthodes supervisées et semi-supervisées, les méthodes non supervisées, ne nécessitent aucune intervention de l'utilisateur.

La classification non-supervisée présente plusieurs avantages importants par rapport à l'approche supervisée :

- i) la non nécessité pour l'utilisateur de préciser à l'avance les classes à discriminer et les échantillons d'apprentissage. L'algorithme de classification détecte automatiquement les classes distinctes de manière objective, ce qui réduit considérablement le risque d'erreur ;
- ii) l'accès à des données d'apprentissage (des relevés de terrain) pour certaines applications est très difficile voir impossible. De plus, il n'est pas rare qu'elles soient imparfaites ou entachées d'erreurs, et que leur utilisation conduise à des aberrations dans les résultats de classification. Par conséquent, les méthodes supervisées ne sont pas appropriées et seules les méthodes non supervisées peuvent apporter une solution objective.

Les méthodes appartenant à la catégorie des méthodes non supervisées peuvent être répertoriées en deux groupes : les méthodes paramétriques dans lesquelles un modèle paramétrique (par exemple, des lois conditionnelles aux classes) est imposé *a priori* ; et les méthodes non paramétriques dans lesquelles aucune hypothèse n'est faite concernant ces lois. Au sein de ce second groupe, les méthodes s'appuient sur des mesures directes de similarité (ou dissimilarité) entre individus pour leur classification.

Les méthodes du premier groupe présentent plusieurs inconvénients en comparaison avec celles du second groupe. Parmi ces inconvénients on peut citer principalement les erreurs d'estimation du modèle du mélange ce qui peut fausser la classification, et plus particulièrement dans le cas des données en grande dimension.

Ce chapitre est donc dédié à un état de l'art sur les méthodes de classification en s'intéressant dans un premier temps aux principales approches de classification non paramétriques semi-supervisées, c'est-à-dire des méthodes qui ne nécessitent pas d'apprentissage, mais un minimum de connaissance *a priori* comme le nombre de classes. Ces méthodes sont : C-moyennes [6], C-moyennes floues [7], ISODATA [8] et SOM [82], [83]. Ensuite, nous présentons deux approches complètement non supervisées et non paramétriques : LBG (Linde, Buzo et Gray) [84] et propagation d'affinité AP [5]. Après une analyse approfondie de l'ensemble de ces méthodes, nous mettons en lumière leurs avantages et inconvénients afin de dégager une nouvelle approche de classification.

## 3.2. La classification

La classification est un procédé algorithmique visant à diviser un ensemble d'individus en un nombre  $N_c$  de classes, selon un critère de similarité dépendant du contexte et du résultat désiré. En imagerie, l'objectif est de produire des partitions mettant en valeur de manière simple et schématique différentes régions élémentaires d'une scène. Ainsi, une classe représente une famille de pixels ayant des propriétés communes et qui identifie une ou plusieurs zones présentant des caractéristiques physiques (type de sol, de végétation, taux d'humidité) ou contextuelles propres. La qualité des résultats obtenus dépendant grandement des classes formées, ces dernières doivent satisfaire aux trois exigences suivantes :

- Exhaustivité : tous les pixels de l'image doivent pouvoir être associés à une classe.

$$\forall x_i \in X, \exists C_j \text{ tel que } x_i \in C_j$$

- Séparabilité : les classes doivent être suffisamment différenciables pour qu'un pixel ne puisse être associé qu'à une seule classe.

$$\text{Soit } C = \{C_1, \dots, C_{N_c}\}, \bigcup_{i=1..N_c} C_i = X \text{ et } \bigcap_{i=1..N_c} C_i = \emptyset$$

- Pertinence : l'association d'un pixel à une classe doit être effectuée suivant un critère objectif à définir et indépendant de l'utilisateur.

Ces trois exigences doivent être vérifiées simultanément pour aboutir à un résultat de classification correct.

Nous présentons dans la section suivante les principales méthodes de classification semi-supervisées et non supervisées.

### 3.2.1. Méthodes de classification semi-supervisées

Une méthode de classification semi-supervisée nécessite un nombre minimum de paramètres. De nombreuses méthodes ont été proposées parmi lesquelles on trouve : trois versions des C-moyennes qui nécessitent au minimum la connaissance *a priori* du nombre de classes (C-moyennes de base [6], la version floue des C-moyennes (FCM) [7] et les C-moyennes possibilistes (PCM) [85]) ; ISODATA [8] qui exige les connaissances *a priori* du nombre de classes, du nombre d'itérations, du nombre d'individus minimal dans une classe, des seuils minimum et maximum d'agrégation, et du nombre maximum de paires de classes pouvant être fusionnées ; et, enfin la carte auto-organisatrice de Kohonen ou Self Organisation Map (SOM) [82] dont le paramètre d'entrée est le nombre de classes. Une

version non supervisée de cet algorithme existe [86] mais elle réduit de manière importante le nombre de classes, ce qui se traduit par un fort taux d'erreur de classification.

#### *3.2.1.1. La famille des C-Moyennes*

L'algorithme des C-Moyennes, proposée en 1967 par McQueen [6], est probablement l'un des algorithmes de classification semi-supervisé les plus connus [87]. Basé sur la minimisation d'une fonction de coût, il est relativement simple et permet d'obtenir de bonnes performances grâce à certaines limitations qui seront discutées plus loin.

##### **Entrée :**

- un tableau croisant l'ensemble  $X$  des individus  $x_i$  à classer et l'ensemble des attributs ( $N \times Na$ ),
- le nombre de classes  $Nc$ .

**Sortie :** Partition de l'ensemble  $X$  en  $Nc$  classes.

##### Algorithme :

- Normaliser les données,
- Sélectionner aléatoirement  $Nc$  individus et les considérer comme les centres de gravité initiaux des classes,
- Tant que le résultat de partitionnement change
  - Affecter chaque individu à la classe  $C_k$  dont le centre de gravité  $g_k$  lui est le plus proche.
  - Calculer les nouveaux centres de gravité des classes en cas de réaffectation.

Fin tant que.

Dans l'algorithme C-moyennes de base, l'appartenance d'un individu à une classe est représentée par une variable binaire. Dans le cas où la décision est non binaire, un degré d'appartenance peut être défini. Ainsi, un élément peut appartenir à plusieurs classes à la fois, avec des degrés d'appartenance dont les valeurs sont comprises entre 0 et 1. Cette notion convient parfaitement à la classification pour modéliser l'appartenance relative d'un point à une classe. On notera ainsi  $D_k(x_i)$  le degré d'appartenance de l'individu  $x_i$  à la classe  $C_k$ . On

distingue trois types d'algorithmes suivant les contraintes imposées sur les degrés d'appartenance :

- Les C-moyennes ("Hard C-Means" ou HCM) [88];

$$\forall x_i \in X : D_k(x_i) \in \{0,1\}, \sum_{k=1}^{N_c} D_k(x_i) = 1 \quad (3.1)$$

- Les C-moyennes floues ("Fuzzy C-Means" ou FCM) [88];

$$\forall x_i \in X : D_k(x_i) \in [0,1], \sum_{k=1}^{N_c} D_k(x_i) = 1 \quad (3.2)$$

- Les C-moyennes possibilistes ("Possibilistic C-means" ou PCM) [85]

$$\forall x_i \in X : D_k(x_i) \in [0,1], 0 < \sum_{k=1}^{N_c} D_k(x_i) < N_c \quad (3.3)$$

Les trois méthodes sont fondées sur la minimisation d'une fonctionnelle commune :

$$J(D, P^r, \omega) = \sum_{k=1}^{N_c} \sum_{i=1}^N D_k(x_i)^m d_{i,k}^2 + \sum_{k=1}^{N_c} \omega_k \sum_{i=1}^N (1 - D_k(x_i))^m \quad (3.4)$$

$D$  est une matrice de taille  $N \times N_c$ , dont les éléments sont les  $D_k(x_i)$ .

$P^r = \{P^r_1, \dots, P^r_{N_c}\}$  est l'ensemble des prototypes des  $N_c$  classes dont la distance aux  $x_i$  est notée  $d_{i,k}$ .

$\omega = \{\omega_1, \dots, \omega_{N_c}\}$  est l'ensemble des termes de pénalité des données atypiques associés à chacune des classes dans le PCM. Ces termes sont identiquement nuls dans le cas de HCM et de FCM.

$m$ , appelé coefficient de flou, est un paramètre de l'algorithme contrôlant la quantité de flou dans la partition ( $m > 1$ ).

Le principe de cette famille de méthodes est de minimiser itérativement la fonctionnelle  $J$  en alternant une mise à jour de  $D$  et de  $P$ .

Le Tableau 2 récapitule les équations de mises à jour des éléments de  $D$ . Pour les méthodes FCM et PCM, les degrés d'appartenance sont, en pratique, proches pour  $m$  supérieur à 20. A l'inverse, si  $m$  tend vers 1, la matrice de partition floue devient une partition stricte. Le choix de  $m$  reste ainsi un problème ouvert. Les utilisateurs de cet algorithme choisissent généralement  $m$  entre 1 et 2 pour des raisons d'efficacité.



Tableau 2: Mise à jour des degrés d'appartenance dans la famille des C-moyennes

HCM	$D_k(x_i) = \begin{cases} 1 & \text{si } d_{i,k} < d_{i,j}, k \neq j \\ 0 & \text{si } k = j \end{cases}$	(3.5)
FCM	$D_k(x_i) = \left[ \sum_{j=1}^{N_c} \left( \frac{d_{i,k}}{d_{i,j}} \right)^{\frac{2}{m-1}} \right]^{-1}$	(3.6)
PCM	$D_k(x_i) = \left[ 1 + \left( \frac{d_{i,k}^2}{\omega_k} \right)^{\frac{1}{m-1}} \right]^{-1}$	(3.7)

De nombreuses adaptations de la famille C-Moyennes ont été proposées [89]–[92], reposant toutes sur le calcul de centres représentatifs des différentes classes. Ce type d’algorithmes souffre de deux principaux inconvénients. Premièrement, le nombre de centres  $N_c$  (nombre de classes) doit être défini *a priori*. Cela nécessite soit de connaître à l’avance le nombre de classes que l’on souhaite obtenir, ce qui est rarement le cas si on veut découvrir des classes naturelles dans les données, soit d’exécuter l’algorithme un grand nombre de fois en variant les valeurs de centres et de choisir la meilleure partition selon un indice de qualité à définir, ce qui est particulièrement coûteux en temps de calcul. Deuxièmement, ces algorithmes ne peuvent pas discriminer des classes contenant des données plus proches du centre d’une autre classe que du centre de leur propre classe. Malheureusement cela arrive fréquemment dans les classes naturelles [93], par exemple lorsque les tailles des classes présentent une grande disparité ou lorsque les classes ne sont pas hyper-sphériques. Enfin, les résultats sont sensibles au paramètre de flou qui peut varier d’une application à l’autre [88].

### 3.2.1.2 La méthode ISODATA

La méthode ISODATA est un algorithme itératif, comme les C-moyennes. L’hypothèse fondamentale de l’algorithme ISODATA est que la distribution de chaque classe suit une loi décroissante à partir de sa moyenne  $\mu_k$ . L’affectation optimale d’un individu  $x_i$  à une classe peut donc se faire en sélectionnant la classe  $C_k$  dont le centre de gravité  $g_k$  est le plus proche de  $x_i$  selon une métrique donnée.

**Entrée :**

- un tableau croisant l'ensemble  $X$  des individus  $x_i$  à classer et l'ensemble des attributs ( $N \times N_a$ ),
- Connaissances *a priori* à fournir par l'utilisateur :
  - le nombre minimum  $N_{min}$  et maximum  $N_{max}$  de classes  $N_c$ ,
  - le nombre maximum d'itérations,
  - le nombre minimum de pixels  $N_{min}$  par classe,
  - le seuil maximum de la variance des classes,
  - la distance minimale entre les centres de gravité des classes,
  - le nombre maximum de paires de classes pouvant être fusionnées.

**Sortie :** Partition de l'ensemble  $X$  en  $N_c$  classes.

*Algorithme :*

1. Normaliser les données,
2. Affecter aléatoirement les  $N$  individus à  $N_{min}$  classes,
3.  $t = 0$ ,
4. Tant que le résultat de partitionnement change OU  $t < t_{max}$ 
  - $t = t + 1$ ,
  - Calculer les centres de gravité des classes  $g_i$ ,
  - Affecter chaque individu à la classe dont le centre est le plus proche,
  - Éliminer les classes comportant moins de  $N_{min}$  individus, et affecter les individus orphelins aux classes dont les centres sont les plus proches,
  - Recalculer les centres des classes,
  - Si deux classes sont suffisamment proches suivant la distance L1, les rassembler,
  - Si la valeur de la variance intra-classe est trop importante, partitionner la classe en deux classes,
  - Recalculer les centres des classes

Fin tant que.

Cet algorithme fournit une approximation des moyennes des distributions de probabilité par classe, selon le critère du maximum de vraisemblance. Il donne de bons résultats dans le cas de distributions normales [94]. Pour les données réelles, cet algorithme est

moins performant ; de plus, il nécessite la connaissance *a priori* de plusieurs paramètres dont le réglage est difficile à optimiser : le nombre minimum et maximum de classes, le nombre maximum d'itérations, le nombre minimum de pixels par classe, le seuil maximum de la variance des classes, le seuil de la distance minimale entre les centres de gravité des classes et le nombre maximum de paires de classes pouvant être fusionnées.

### 3.2.1.3 La carte auto-organisatrice

La carte auto-organisatrice ou Self Organizing Map (SOM) [82], [83] encore appelée algorithme de Kohonen, est un algorithme stochastique de classement des données, qui respecte la topologie de l'espace des observations en intégrant une notion de voisinage entre les classes. Cet algorithme est souvent utilisé pour la réduction de la dimensionnalité et la visualisation des données [95]–[97].

L'algorithme comporte deux étapes principales : l'extraction des caractéristiques et la classification fondée sur un réseau de neurones (voir structure en Figure 8). Un neurone est une unité de traitement de l'information. Les neurones sont connectés par des synapses, chacune d'entre elles caractérisées par un poids. Plus précisément, un signal  $x_j$  à l'entrée  $j$  du synapse reliée au neurone  $k$  est multiplié par le poids synaptique  $w_{kj}$ . Un réseau de neurone est un ensemble de neurones organisés sous la forme de couches. Si la couche d'entrée possède  $N$  unités et la couche de sortie possède  $M$  unités, chaque unité dans la couche de sortie possède  $N$  poids associés aux connexions qui proviennent de la couche d'entrée, de sorte que l'ensemble des poids de neurones est organisé sous la forme d'un réseau à deux dimensions  $W_{M \times N}$ .

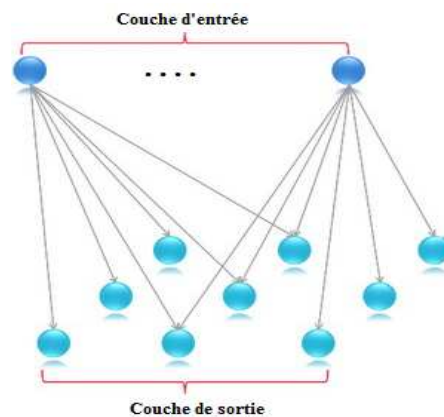


Figure 8: Topologie de l'algorithme SOM

Les étapes de l'algorithme SOM semi-supervisé sont les suivantes :

**Entrée :**

- un tableau croisant l'ensemble  $X$  des individus  $x_i$  à classer et l'ensemble des attributs ( $N \times Na$ ),
- le nombre de classes  $Nc$ .

**Sortie :** Partition de l'ensemble  $X$  en  $Nc$  classes.

*Algorithme :*

1. Normaliser les données,
2.  $t = 0$ ,
3. Initialiser aléatoirement les  $Nc$  prototypes  $P^r_i = (P^r_{i1}, \dots, P^r_{iNa})$  pour chaque neurone  $i$ ,  
 $1 \leq i \leq Nc$ .
4. Tant que le résultat de partitionnement change

- $t = t + 1$ ,

Étape de compétition :

- Echantillonnage : choisir aléatoirement un individu  $x_k$  représenté par son vecteur d'attributs  $a_k$ ,
- Correspondance et similarité : parmi les  $Nc$  neurones, choisir le meilleur,  $v^*(x_k)$ , pour représenter cet individu avec :

$$v^*(x_k) = \underset{1 \leq i \leq Nc}{\text{Arg min}} \left\{ d(a_k, P^r_i) \right\} \quad (3.8)$$

où  $d$  est la distance euclidienne

Étape d'adaptation :

- Apprentissage : mettre à jour les prototypes  $P^r_j$  de chaque neurone  $j$  selon la règle :

$$P^r_j(t) = P^r_j(t-1) - \varepsilon(t) K_{jv^*(x_k)} (P^r_j(t-1) - a_k) \quad (3.9)$$

où

$K_{jv^*(x_k)}$  est une fonction de voisinage centrée autour des neurones sélectionnés définie par :

$$K_{j_{v^*(x_k)}} = t \left[ \frac{e^{-d(v^*(x_k), v(x_k))^2}}{d_{r1}} \right] \text{ avec } t \geq 1 \quad (3.10)$$

$\varepsilon(t)$  est le taux d'apprentissage, dépendant de la valeur de l'itération  $t$  avec  $\varepsilon(t) = \frac{1}{t}$  pour  $t \neq 0$ .

et  $d_{r1}$  est la taille du voisinage à la première itération.

Fin tant que.

La performance du réseau de neurones SOM dépend de plusieurs paramètres de réglage :

- Nombre de neurones de sortie : le nombre idéal de neurones de sortie doit être égal au nombre de classes souhaité, associant exactement un neurone avec une classe.
- Initialisation des poids : les poids sont initialisés de manière aléatoire.
- Choix de la fonction de voisinage (p. ex. fonction gaussienne)
- Distance de référence  $d_r$  qui détermine la taille du voisinage.

Cet algorithme est souvent utilisé comme première étape d'analyse [95]–[97] ; de nombreux travaux ont montré son efficacité en termes de pré-classification. En effet, SOM estime des référents représentant des sous classes, et les partitions associées à chaque référent sont ensuite utilisées en entrée d'une méthode de classification traditionnelle pour former la classification finale. D'après ces études, l'introduction du SOM comme étape préliminaire avant celle de la classification pour réduire la dimension des données, contribue à un meilleur 2partitionnement des images.

#### 3.2.1.4 L'immunité artificielle

Il a été proposé d'utiliser le concept de système immunitaire artificiel pour la classification, avec notamment une application aux images multi-composantes. Ce concept, proposé depuis la fin des années 1990 [141], a été utilisé dans de nombreux domaines comme la reconnaissance de formes [142], la classification non supervisée [143] et l'optimisation [143]. La méthode proposée, nommée UAIC (*Unsupervised Artificial Immune Classifier*), est en réalité une méthode semi-supervisée, qui

requiert comme préalable le nombre de classes. Elle est basée sur la mutation de populations d'anticorps (*antibodies*) engendrés par clonage à partir de cellules mémoire (*memory cells*), réagissant à la présentation d'antigènes (*antigens*), et représentant l'ensemble des centres de classes. Dans la mise en œuvre proposée, les antigènes correspondent aux données (vecteurs d'attributs correspondant aux pixels de l'image) non encore rencontrées au cours du traitement.

La sélection clonale des cellules de types lymphocytes *B* dans le système immunitaire humain est le mécanisme fondamental duquel s'inspire l'UAIC. Le Tableau 3 résume l'analogie entre le système immunitaire humain et UAIC.

Tableau 3 : Correspondance entre le système immunitaire humain et UAIC

Système immunitaire	UAIC
Antigènes	vecteurs d'attributs des pixels de l'image
Anticorps	centres des classes/vecteurs des caractéristiques
Espace des formes	ensemble des valeurs possibles des vecteurs/pixels
Expansion clonale	reproduction des centres des classes qui sont bien assortis avec des antigènes
Maturation d'affinité	proportion entre la mutation des centres de classes et la suppression des centres de classes les plus faiblement stimulés
Mémoire immunitaire	mise en mémoire des centres de classes mutés
Méta dynamique	création et suppression continue des centres de classes

La mise en œuvre d'UAIC comprend deux étapes : tout d'abord, le premier regroupement des centres est effectué par un choix aléatoire depuis les données en entrée. Ensuite, l'étape de classification est effectuée. Cette méthode assigne chaque pixel à la classe qui maximise la stimulation entre l'antigène et l'anticorps. Par la suite, basée sur la classe, la population d'anticorps évolue et le groupe de cellules mémoire est mis à jour par des algorithmes immunitaires jusqu'à ce que le critère d'arrêt soit rempli.

L'algorithme UAIC est présenté en trois étapes : initialisation, classification avec UAIC et condition d'arrêt.

Soient  $AB$  l'ensemble des anticorps,  $ab$  un seul anticorps particulier ( $ab \in AB$ ),  $MC$  l'ensemble des cellules mémoire et  $mc$  un individu de cet ensemble

- **Initialisation**

Cette étape peut être considérée comme un prétraitement des données. UAIC applique l'approche Kaufman (KA) [98] à la population initiale de cellules mémoire  $MC$ .

1 : Sélection de la première  $MC$  (cellule mémoire) la plus centralisée.

2 : pour tout anticorps non sélectionné  $ab_i$  faire

- pour tout anticorps non sélectionné  $w_i$  faire

calculer  $T_{ji} = \max(D_j - d_{ji}, 0)$  avec  $d_{ji} = \text{dist}(ab_j, ab_i)$  et  $D_j = \min_s (d_{sj})$  ( $s$  est l'une des  $MC$  sélectionnée),

- calculer le gain de la sélection de  $ab_i$  avec  $\sum_j T_{ji}$ .

3 : Sélection des instances  $ab_i$  non encore sélectionnées qui maximisent  $\sum_j T_{ji}$

4 : Si le nombre de  $MC$  sélectionnées est  $T$  alors Stop ; sinon Revenir à 2,

5 : Pour terminer le regroupement, assigner chaque anticorps non sélectionné  $Ab_{\{r\}}$  au groupe représenté par la cellule mémoire la plus proche aux cellules initiales  $Ab_{\{r\}}$  suivant l'indice :

$$S_{SAM}(x_i, x_j) = \cos^{-1} \left[ \frac{\langle a_i, a_j \rangle}{\|a_i\| \times \|a_j\|} \right] \quad (3.11)$$

qui mesure la distance entre les deux individus  $x_i$  et  $x_j$ .

- **Classification utilisant UAIC**

Une fois l'initialisation terminée, l'étape suivante est une étape de classification itérative. Pour chaque itération, l'algorithme effectue les étapes suivantes pour former chaque antigène ( $ag$ ).

1 : Attribuer l'antigène  $ag$  à la  $k$  ième classe : Pour chaque  $ag$  dans l'image, attribuer cet antigène à l'une des classes, où la classe est supposée être la  $k$  ème classe. Étant donné un antigène spécifique, il faut trouver la cellule-mémoire  $mc$  qui maximise l'affinité :

$$mc = \arg \max_{mc \in MC} affinity(ag, mc) \quad (3.12)$$

$$\text{avec } affinity = \exp\left(-\frac{dist(x, y)}{2u^2}\right) \quad (3.13)$$

où  $u$  est le rayon d'influence de  $AB$ .

- 2 : Évolution de la population d'anticorps  $AB^k$  : Après avoir assigné les antigènes aux classes, l'évolution de la population d'anticorps  $AB^k$  et de la cellule mémoire  $MC^k$  sont effectuées.
- 3 : Mise à jour de la mémoire  $MC^k$  : L'étape finale dans le processus de formation est l'introduction potentielle des nouvelles cellules mémoire développées  $mc$  dans l'ensemble des cellules mémoire  $MC$  existantes.
- 4 : Agrégation des cellules mémoire contenant la même information afin de limiter la croissance de population.

- **Condition d'arrêt**

La condition d'arrêt est différente suivant les applications. Une solution consiste à définir un nombre fixe d'itérations comme condition d'arrêt. Une autre solution consiste à fixer un seuil. En pratique, un seuil est fixé suivant la proportion de pixels de chaque classe qui changent de classe au cours des itérations de l'algorithme.

Enfin, UAIC renvoie le résultat de classification de l'image.

Appliquée sur deux images multispectrales (Landsat TM, MODIS) et une image hyperspectrale (PHI, 80 bandes), cette méthode a montré des résultats supérieurs aux méthodes k-means, ISODATA, FCM et SOM avec des taux de bonne classification supérieurs à 80%, soit en moyenne 10% de plus que la moins efficace des méthodes comparées, en l'occurrence la méthode  $k$ -means.

### 3.2.2. Méthodes de classification non supervisée

Pour la classification non supervisée, nous ne disposons d'aucune autre information que les données elles mêmes. Ainsi, ces méthodes ne sont pas sujettes au choix par l'utilisateur de paramètres qui peuvent grandement affecter le résultat. Nous présentons ici



quatre approches qui répondent à ces exigences d'indépendance à toute information *a priori*. Nous verrons ainsi les méthodes génétiques, l'approche FCM non supervisée, la méthode MLBG et enfin, la méthode de propagation d'affinité.

### 3.2.2.1 Les méthodes génétiques

Les algorithmes génétiques (AG), présentés par John Holland en 1975 [99], [100] sont des algorithmes inspirés du processus d'évolution naturelle. Dans de nombreux domaines applicatifs, les algorithmes génétiques ont été appliqués avec succès [101]. En 1975, De Jong [102] a effectué un grand nombre de tests pour étudier l'effet du changement des paramètres de contrôle sur les performances de l'AG. Lors de ces tests, différentes valeurs ont été analysées, telles que la taille de la population, la probabilité de croisement et la probabilité de mutation [100]. En 2001, Bandyopadhyay et Maulik [103] ont appliqué l'AG pour regrouper différents ensembles de données et les fusionner.

- **L'algorithme génétique**

L'approche génétique est une méthode adaptée pour résoudre une très large gamme de problèmes. Récemment, AG a été appliquée avec succès pour l'optimisation des problèmes de classification non supervisée [104], [105].

Les principales étapes sont :

1 : Initialisation de la population (Chromosomes) et définition de la fonction d'aptitude ou de « *fitness* ».

Une population est l'ensemble des chromosomes. La taille typique de la population peut aller de 20 à 1000 [99].

Pour chaque chromosome de la population, un paramètre  $K_i$  correspondant à la longueur du chromosome (équivalent au nombre des classes dans le problème de classification) est généré aléatoirement dans l'intervalle  $[K_{min}, K_{max}]$ , où  $K_{min}=2$  généralement et  $K_{max}$  représente la longueur maximale des chromosomes. Ce chromosome est supposé coder les  $K_i$  centres de classes. Pour l'initialisation de ces centres, les  $K_i$  individus sont choisis aléatoirement à partir de l'ensemble de données. Ces individus sont distribués de manière aléatoire dans le chromosome.

Une fonction d'aptitude doit être définie pour mesurer l'adaptabilité de chaque chromosome de la population. L'aptitude peut être calculée à l'aide d'un indice comme celui de Davies-Bouldin [144].

## 2 : Sélection et mutation de la population courante.

Au cours de la mutation, tous les chromosomes de la population sont vérifiés unité par unité, et selon une probabilité prédéfinie. Toutes les valeurs d'une unité spécifique peuvent être changées au hasard.

## 3 : Croisement:

Le but de l'opération de croisement est de créer deux nouveaux chromosomes à partir de deux chromosomes existants choisis au hasard parmi la population actuelle.

## 4 : Évaluation et arrêt

Basé sur le croisement et la mutation des chromosomes, une fois initialisé, l'AG évolue itérativement d'une génération à l'autre. La population évolue au fil des générations dans le but de maximiser la valeur de la fonction d'aptitude.

L'algorithme continue tant que le critère d'arrêt n'est pas satisfait.

L'algorithme génétique permet la résolution d'un problème de grande complexité par l'examen simultané de solutions multiples. Cependant, il présente quelques difficultés de mise en œuvre, comme le choix de la fonction d'aptitude dont dépend la qualité des résultats, et la nécessité de disposer d'une population importante de chromosomes.

### • **Algorithme Génétique Hybride**

Dans [56] est décrit un algorithme génétique hybride incorporant, dans un schéma multi-résolution, l'algorithme génétique traditionnel et la méthode de classification *k*-means. Cette méthode est une autre variante de l'AG dans laquelle l'opérateur de croisement est remplacé par la classification C-moyennes, et les autres opérateurs demeurent inchangés. Ce remplacement est justifié du fait que l'opérateur de sélection de l'algorithme génétique prélève des solutions en accord avec les valeurs d'aptitude du chromosome (information *globale*), au lieu d'utiliser l'interaction locale entre les gènes.

Dans cette méthode, une structure *quad-tree* est d'abord construite, et l'image est partitionnée en blocs à différents niveaux de résolution. Des attributs de texture sont ensuite extraits de chaque bloc (la valeur moyenne des niveaux de gris est le seul attribut de texture utilisé).

L'image entière est alors combinée en tant que chromosome/solution, chaque bloc étant vu comme un gène. Ainsi, chaque gène sera mis en relation avec quatre voisins immédiats, sauf ceux situés sur les bords de l'image/chromosome.

L'algorithme génétique hybride est alors exécuté pour partitionner l'image en apposant un label de classe de texture optimal pour chaque gène. Une fois la convergence atteinte à un niveau donné du *quad-tree*, l'allèle de chaque gène (résultat de partitionnement) du niveau courant est propagé comme allèle initial des gènes descendants au niveau inférieur. L'algorithme est stoppé à chaque niveau de résolution lorsqu'un critère d'arrêt est atteint.

Les différentes étapes de cet algorithme pour chaque niveau de résolution peuvent être résumées comme suit :

1 : Extraction des attributs à partir de chaque bloc,

2 : Partitionnement du niveau courant :

- Initialiser la population
- Faire
  - Appliquer la classification *C-moyennes*
  - Appliquer la mutation
  - Évaluer la fonction d'aptitude
  - Sélectionner les chromosomes en fonction de leur aptitude pour la génération suivante

Tant qu'il y a évolution de la population

3 : Propagation du résultat de partitionnement au niveau inférieur,

4 : Retour à l'étape 1.

L'application de cet algorithme sur des images monocomposantes montre que le taux d'erreur de classification décroît lorsque le niveau de résolution augmente. Cette méthode a cependant le même inconvénient que l'algorithme *C-moyennes* (sensibilité au choix aléatoire des centres des classes).

#### *3.2.2.2 FCM non supervisé*

Dans [106], une version non supervisée du FCM, adaptée aux problèmes de reconnaissance de formes multi-classes en hautes dimensions, est présentée dans l'objectif d'accroître la précision et la stabilité du FCM standard. Elle s'appuie sur deux concepts : la pondération non supervisée des centres de classes à partir de l'extraction non paramétrique d'attributs pondérés, et l'extraction d'attributs par analyse discriminante. L'avantage de cette

méthode réside dans ses caractéristiques non supervisée et non paramétrique, grâce auxquelles le système détermine automatiquement le nombre de classes ; de plus, elle est plus robuste que l'algorithme FCM classique. Toutefois, elle n'est pas complètement stable car, suivant la complexité des images traitées, le niveau de variabilité peut être élevé.

### 3.2.2.3 La méthode MLBG (Modified LBG)

La méthode MLBG proposée par Rosenberger et Chehdi [41] est une optimisation de la méthode semi-supervisée LBG [84]. La méthode MLBG ne nécessite pas la connaissance du nombre de classes mais seulement d'un majorant de celui-ci. Elle est caractérisée par les propriétés suivantes :

- évaluation et possibilité de remise en cause d'un résultat intermédiaire,
- correction d'un résultat courant en exploitant les résultats antérieurs,
- évaluation de la solution sans connaissance de vérité de terrain.

Pour déterminer la meilleure partition de  $X$ , quatre étapes essentielles sont nécessaires.

**Entrée :** un tableau croisant l'ensemble  $X$  des individus  $x_i$  à classer et l'ensemble des attributs ( $N \times N_a$ )

**Sortie :** Partition de l'ensemble  $X$  en  $N_c$  classes.

*Algorithme :*

1. Normalisation des données
2. Choix de la classe à subdiviser : à l'état initial ( $k = 1$ ), la classe à subdiviser est l'ensemble  $X$ . A chaque itération ( $k > 1$ ), la classe qui maximise une fonction de dispersion est partitionnée,
3. Choix de noyaux de division d'une classe, composés du barycentre de la classe à partitionner et du point le plus distant du barycentre appartenant à la même classe,
4. Partition en  $k+1$  classes par la méthode des C-moyennes,
5. Evaluation de la partition de l'ensemble  $X$  suivant le critère de dispersion intra-classe ( $Dis$ ) Eq. (1. 2). Si ces critères sont satisfaits (la partition de l'ensemble  $X$  à  $k + 1$  classes est valide), alors l'étape 1 est relancée pour essayer de créer  $k + 2$  classes. Dans le cas contraire, le choix des noyaux est remis en cause et

l'étape 2 est de nouveau relancée en intégrant les échecs précédents. Si aucun noyau n'engendre une partition valide, le choix de la classe à subdiviser est remis en cause et l'étape 1 est relancée. Si aucune classe ne génère une nouvelle partition valide alors le nombre de classes et la partition obtenus sont considérés optimaux.

La dispersion intra-classe est donnée par :

$$Dis(C_i) = \frac{1}{card(C_i)} \sum_{j=1}^{N_{C_i}} d(g_i, a_i^j) \quad (3.14)$$

où  $g_i$  est le centre de gravité de la classe  $C_i$  et  $a_i^j$  est l'attribut du  $j$ ème élément de la classe  $C_i$ .

Le point sensible de cette méthode réside dans l'initialisation des noyaux. En effet, le choix des noyaux des classes influe sur la convergence et la qualité du résultat.

#### 3.2.2.4 La propagation d'affinité

Récemment, Frey et Dueck [5] ont proposé une nouvelle approche de classification qui peut être utilisée de manière totalement non supervisée ou semi-supervisée : la propagation d'affinité. Cette méthode a connu un vif succès dans plusieurs domaines applicatifs pour les trois principales raisons suivantes : *i)* elle est efficace lorsque le nombre de classes est connu, *ii)* elle peut être utilisée suivant deux modes : non supervisé ou semi-supervisé, *iii)* elle est complètement déterministe.

L'AP a été appliquée dans différents domaines telles que la télédétection [9], [107], la gestion des données multimédia [108] et la reconnaissance des formes (catégorisation des images [109], la reconnaissance des chiffres manuscrits [110], la reconnaissance de la parole [111], les requêtes visuelles [112], la caractérisation des protéines [113] et la génomique [145]).

#### • AP pour la réduction de la dimensionnalité

Dans le domaine de la télédétection, l'AP a été utilisée principalement pour réduire le nombre d'attributs ou la sélection de bandes spectrales pour les images hyperspectrales. Ceci peut s'expliquer par le fait que le nombre de bandes spectrales (considérées ici comme des individus) dans l'imagerie hyperspectrale est limité à quelques centaines, ce qui peut être géré facilement par l'AP malgré sa complexité quadratique. La première application de la méthode AP aux images hyperspectrales pour la sélection des bandes spectrales a été proposée par Jia

et Qian dans [9]. Dans [114], Qian propose une amélioration de la méthode présentée dans [9], par l'introduction de la divergence de Kullback-Leibler [115] comme mesure de similarité entre les bandes spectrales, ainsi que le kurtosis de l'ensemble des bandes comme paramètre de préférence pour la classification des images. La comparaison pour la sélection de bandes de l'AP avec l'algorithme basé sur la divergence de l'information (ID) et celui de l'analyse en composantes principales par maximum de vraisemblance (MV PCA) a été réalisée sur trois images hyperspectrales issues de trois capteurs hyperspectraux différents (AVIRIS Indian Pines, HYDICE Washington D.C. Mall et ROSIS Pavia). Les résultats obtenus, donnés dans le Tableau 4(a), montrent que l'AP permet de mieux sélectionner les bandes les plus représentatives des données. Il est à noter que le classifieur supervisé SVM est utilisé dans l'étape de classification des données après réduction.

Dans [116], Jia et al. introduisent une étape d'extraction d'attributs utilisant différents types d'ondelettes (Haar, Daubechies et Coiflet) à différents niveaux. Ensuite, ils sélectionnent avec l'AP les attributs les plus représentatifs pour les intégrer dans le classifieur KNN. L'application de cette approche sur des images hyperspectrales AVIRIS montre qu'en procédant à la sélection d'un certain nombre d'ondelettes par l'AP, on obtient de meilleurs résultats de classification qu'avec l'ensemble des bandes utilisant directement les signatures spectrales des pixels ou un sous-ensemble de bandes sélectionnées par l'AP. Pour 13 attributs sélectionnés par l'AP parmi les 220 attributs (un coefficient d'approximation au premier niveau par bande), les ondelettes de Haar et de Daubechies donnent respectivement 83.5%, et 82.5% de taux de bonne classification. Quant à la sélection de bandes spectrales de l'image par l'AP, elle ne conduit qu'à un taux de bonne classification de 81%.

Jia et al. ont également développé une méthode non supervisée de sélection de bandes pour les images hyperspectrales sans suppression manuelle de bandes [10]. Cette approche utilise d'abord les ondelettes pour retirer automatiquement le bruit spatial sur toutes les données hyperspectrales, puis applique l'AP afin de choisir les bandes les plus représentatives et enfin utilise un algorithme de classification (KNN ou SVM) pour évaluer la pertinence des bandes sélectionnées. L'application sur l'image hyperspectrale AVIRIS Indian Pines montre que l'utilisation des attributs après réduction par AP donne de meilleurs résultats de classification que l'utilisation de l'ensemble des bandes (+5% de différence). Pour l'image multispectrale Washington D. C. Mall les résultats sont presque similaires à ceux utilisant l'ensemble des bandes (+1%). L'AP donne un meilleur résultat que l'algorithme basé sur l'information mutuelle (MI) pour la sélection de bandes (+2% de différence par rapport à la

classification avec toutes les bandes) pour AVIRIS et -3% de différence pour Washington D. C. Mall.

Dans [117], une méthode semi-supervisée basée sur l'application de l'AP avec différentes métriques (corrélation et séparabilité entre les bandes) a été proposée. Cette méthode nommée FM-AP (Feature Metric AP) a été appliquée avec succès sur des images hyperspectrales HYPERION (250 x 566 pixels x 224 bandes) et AVIRIS Indian Pines (145 x 145 pixels x 220 bandes) pour la sélection de bandes spectrales. Les auteurs ont comparé leur méthode de réduction à l'analyse en composantes principales par maximum de vraisemblance (MV PCA), à la sélection de bande basée sur le critère de corrélation (PCC Pearson correlation coefficient) et à l'AP standard en utilisant le SVM comme classifieur. Les résultats de classification sont également reproduits dans le Tableau 4(b), et montrent encore la supériorité de la méthode de sélection proposée sur ces deux jeux de données, avec un maximum du taux de bonne classification obtenu pour un nombre de bandes autour de la cinquantaine.

À travers les différentes expérimentations [109], [114], [117] l'AP standard a donné généralement les meilleurs résultats dans la sélection de bandes pour différentes images hyperspectrales (AVIRIS, HYDICE, HYPERION) par rapport à d'autres approches, y compris l'analyse en composantes principales par maximum de vraisemblance (MV PCA), la divergence de l'information (ID) et l'information mutuelle (MI).

L'AP a été utilisée pour la détection des régions d'intérêt et la sélection des paysages représentatifs pour la télédétection, mais seulement sur des images de petites tailles spatiales.

Dans [118], Yang et al. ont proposé de remplacer la distance euclidienne de l'AP standard par une mesure de similarité basée sur les statistiques floues comme entrée de l'algorithme AP (FS-AP). La méthode a été comparée aux C-moyennes, FCM et l'AP standard en utilisant trois types d'images multispectrales de télédétection de petite taille (Landsat ETM+ : 86 x 86 pixels x 6 bandes ; Quickbird : 88 x 88 pixels x 3 bandes et Modis : 90 x 86 pixels x 5 bandes). La méthode FS-AP améliore les résultats de classification par rapport à la version standard de l'AP avec une charge de calcul plus faible et donne de meilleurs résultats que les autres algorithmes cités. Les performances comparées des différentes méthodes sur ces images multispectrales sont données dans le Tableau 5(a), et montrent l'apport de la méthode FS-AP.

Dans [119], Yang et al. proposent une approche semi-supervisée de l'AP basée sur l'apprentissage incrémental-décrémental nommée ID-AP. Elle est appliquée sur les trois types d'images multispectrales utilisées dans [118] pour la classification de la couverture terrestre. Comparée à différents algorithmes de classification semi-supervisée Constrained K-means (CKM), AP semi-supervisée (SAP), et SAP incrémental (IS-AP), l'approche ID-AP donne les meilleurs taux de classification comme le montre le Tableau 5(b). On peut noter que les trois méthodes à base de propagation d'affinité sont meilleures que la méthode CKM dans tous les cas.

D'autres extensions de l'AP ont été proposées dans [117] et [118] et appliquées sur des images multispectrales de petite taille (moins de 100x100 pixels).

Dans [110], une extension du modèle à un seul représentant de l'AP à un modèle multi représentants nommé ME-AP a été développée. Cette approche détermine le nombre de représentants dans chaque classe associés à un représentant global afin de rapprocher les sous classes. Elle a été appliquée sur trois bases de données d'images et deux bases de données de chiffres manuscrits. Les résultats ont montré de meilleures performances que l'AP standard avec un modèle de représentant unique, la version hiérarchique de l'AP [120] et la classification basée sur la théorie spectrale des graphes. L'AP a également été associée à l'apprentissage semi-supervisé nommé SAP (Seeds Affinity Propagation) [121]. Cette méthode est une version semi-supervisée de l'AP qui utilise une mesure de similarité asymétrique et capte l'information structurelle de textes en introduisant une approche d'apprentissage semi-supervisée, où la connaissance de quelques objets étiquetés par rapport à un grand nombre de ceux non marqués est exploitée. La comparaison de l'approche SAP aux C-moyennes et à l'AP standard sur les données de texte (avec des définitions de similarité adaptées) a montré de meilleures performances en termes de précision de la classification.

L'AP a été également récemment utilisée avec succès pour l'identification des paysages représentatifs des écosystèmes forestiers dans dix régions forestières du Canada à partir des images satellite Landsat [122].

La supériorité de l'AP par rapport à d'autres méthodes de classification classiques a été aussi démontrée dans une étude comparative sur la reconnaissance de visage [5], où elle a fourni de meilleurs taux de classification que les C-moyennes, C-centres et la classification ascendante hiérarchique.



Toutes les études citées dans cette section montrent la supériorité de la méthode de propagation d'affinité par rapport aux méthodes classiques, soit pour la réduction d'attributs (ou des bandes spectrales), soit pour la classification des pixels sur des images de petites tailles.

- **AP pour la classification**

Si l'AP a été largement utilisée pour réduire le nombre d'attributs dans le domaine de la télédétection, elle l'a été beaucoup moins pour la classification des pixels. Cela est dû à la complexité quadratique de l'algorithme qui exige le calcul préalable d'une matrice de similarité de taille  $N \times N$ ,  $N$  étant le nombre d'individus.

Malgré ses bonnes performances de classification, l'application de l'AP sur des données réelles de grandes dimensions spatiales comme les images aériennes hyperspectrales reste impossible. Par exemple, sur un PC de 4 Go RAM, 3GHz processeur avec un système d'exploitation 32 bits (environnement Matlab), le nombre d'individus qui peuvent être classés par l'algorithme AP est généralement limité à moins de 3000 pixels ou objets. En outre, on observe expérimentalement que la version non supervisée surestime fortement le nombre de classes.

Tableau 4: Comparaison des méthodes de réduction des bandes des images hyperspectrales et classification avec SVM (ACCR : Taux de bonne classification moyen (en %)).

Qian et al., 2009 [9]	MV-PCA	ID	AP
AVIRIS /Indian Pines	60.13	62.65	<b>82.67</b>
HYDICE/Washington DC Mall	74.96	87.40	<b>89.06</b>
ROSIS/Pavia	99.71	99.77	<b>99.82</b>

(a)

Yang et al., 2013 [117]	MV-PCA	PCC	AP	FM-AP
AVIRIS/Indian Pines	61.08	74.01	74.02	<b>87.15</b>
HYPERION/Xuzhou	75.59	88.83	85.78	<b>90.21</b>

(b)

Tableau 5: Comparaison des méthodes de classification sur des images multispectrales (ACCR en %).

<b>Yang et al., 2010 [118]</b>	C-moyennes	FCM	AP	FS-AP
Landsat-7 ETM/ Haerbin	65.94	70.01	74.10	<b>82.37</b>
Quickbird /Trento	73.20	74.74	76.60	<b>80.22</b>
MODIS / Changchun	77.75	79.35	78.76	<b>82.76</b>

(a)

<b>Yang et al., 2013 [119]</b>	CKM	SAP	IS-AP	ID-AP
Landsat-7 ETM/ Haerbin	78.54	80	87.17	<b>90.37</b>
Quickbird / Trento	81.59	84.66	84.85	<b>88.67</b>
MODIS / Changchun	81.60	84.02	83.62	<b>89.10</b>

(b)

Nous détaillons ci-dessous l'aspect algorithmique de la méthode de propagation d'affinité.

#### **Algorithme de propagation d'affinité :**

La classification par propagation d'affinité requiert d'abord le calcul d'une matrice de similarité notée  $S$ . Les éléments  $S(x_i, x_k)$  de la matrice  $S$  mesurent la similarité entre tous les couples d'individus  $(x_i, x_k)$  pour  $x_i \neq x_k$ . La distance euclidienne négative est souvent utilisée comme mesure de similarité :

$$S(x_i, x_k) = -\|a_i - a_k\|_2, \forall x_i \neq x_k \quad (3.15)$$

avec  $a_i$  le vecteur d'attributs représentant l'individu  $x_i$ .

Les éléments diagonaux  $S(x_k, x_k)$  de la matrice  $S$ , reflétant la pertinence *a priori* du choix de l'individu  $k$  pour servir de représentant et connus sous le nom de *préférence*, ne sont pas calculés de la même manière que les éléments  $S(x_i, x_k)$  pour  $x_i \neq x_k$ . Plus précisément,  $S(x_k, x_k) = p$  (paramètre de préférence) est initialisé à la valeur médiane des éléments de  $S$  pour  $x_i \neq x_k$ .

$$S(x_k, x_k) = p \forall x_k \quad (3.16)$$

AP est connu comme un algorithme de passage de messages. En effet, à chaque itération de l'algorithme, la question suivante est posée : quel individu sera considéré comme représentant ou exemplaire de tous les autres, et quel représentant sera choisi pour chaque individu ? Deux types de messages sont échangés : “*Responsabilité*” et “*Disponibilité*”. Une fois initialisé,

l'algorithme calcule de manière itérative pour chaque individu ses degrés de disponibilité et de responsabilité envers les autres individus. Initialement, tous les individus sont considérés comme représentants. Deux procédures de transmission de messages appelés *responsabilité* et *disponibilité* sont utilisées pour échanger les messages entre l'individu  $x_i$  et un représentant candidat  $x_k$ . La Figure 9 illustre un exemple de la procédure d'échanges de messages de l'AP. La *responsabilité*  $R(x_i, x_k)$  est le message envoyé à partir de  $x_i$  au candidat représentant  $x_k$ , indiquant combien  $x_k$  serait un bon représentant de  $x_i$ . Alternativement, la *disponibilité*  $A(x_i, x_k)$  est le message envoyé du candidat représentant  $x_k$  à l'individu  $x_i$ , indiquant à quel point il serait approprié pour  $x_i$  de choisir  $x_k$  comme son représentant. Cette procédure cherche pour chaque individu le représentant qui maximise la somme des *disponibilités* et des *responsabilités*.

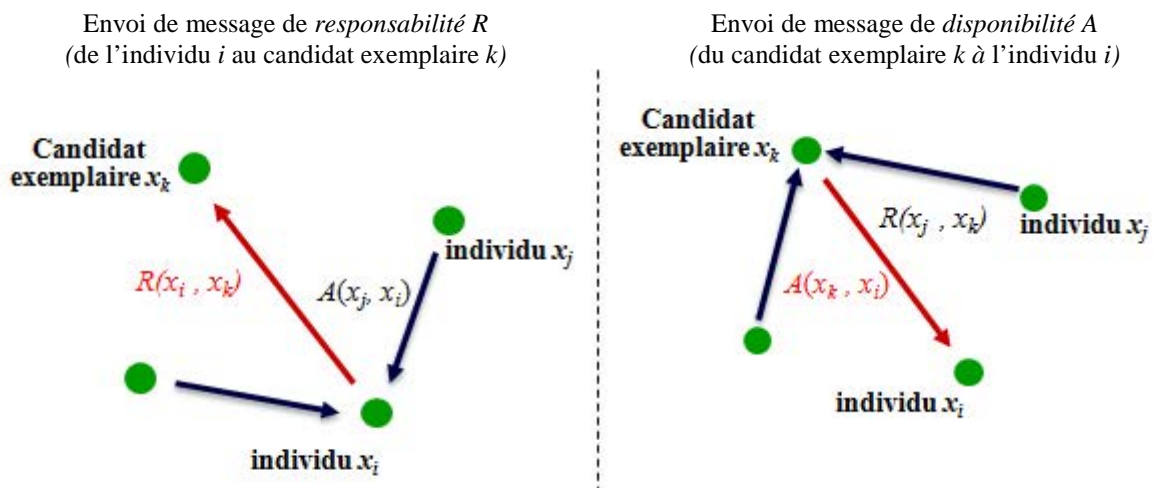


Figure 9: Procédures d'échanges de messages de l'AP

Les quatre principales étapes de l'algorithme sont :

#### Entrée :

- $N$  le nombre d'individus à classer,
- $S$  la matrice de similarité de taille  $N \times N$
- Dans le cas du mode supervisé : le nombre de classes

#### Sortie :

- Résultat de classification sous forme d'un vecteur contenant pour chaque individu l'indice de son exemplaire.

- L'évaluation de la fonction objectif que la propagation d'affinité a maximisée pour chaque individu. Elle indique à quel point les exemplaires choisis sont appropriés pour représenter les données.

### 1. Initialisation :

Soient  $R$  la matrice de responsabilité et  $A$  la matrice de disponibilité. Ces deux matrices, ainsi que la matrice de similarité  $S$ , sont de taille  $N \times N$ .

$$R(x_i, x_k) = 0, A(x_k, x_i) = 0 \quad \forall x_i, x_k \quad (3.17)$$

### 2. Mise à jour de la responsabilité

$$R(x_i, x_k) = S(x_i, x_k) - \max_{j, j \neq k} (A(x_j, x_i) + S(x_i, x_j)) \quad \forall x_i \neq x_k \quad (3.18)$$

### 3. Mise à jour de la disponibilité

$$A(x_k, x_i) = \min\{0, R(x_k, x_i) + \sum_{(j, j \neq k)} \max\{0, R(x_j, x_k)\}\} \quad \forall x_i \neq x_k \quad (3.19)$$

$$A(x_k, x_k) = \sum_{(j, j \neq k)} \max\{0, R(x_j, x_k)\} \quad \forall x_k$$

### 4. Affectation des représentants

$$E^*(x_i) = \arg \max_k \{ (R(x_i, x_k) + A(x_k, x_i)) \}$$

Avec  $E^*(x_i)$  le représentant attribué à l'individu  $x_i$ .

Un facteur d'amortissement  $\alpha \in [0,1]$  est introduit pour régulariser la mise à jour des matrices de responsabilité et de disponibilité au cours des itérations. Pour éviter les oscillations numériques qui se présentent dans certaines circonstances, il est recommandé d'utiliser  $\alpha=0.9$  [120]. Les mises à jour des messages sont réalisées de manière itérative suivant les relations suivantes :

$$\begin{aligned} \hat{R}_{it} &= \alpha \hat{R}_{it-1} + (1-\alpha) R_{it} \\ \hat{A}_{it} &= \alpha \hat{A}_{it-1} + (1-\alpha) A_{it} \end{aligned} \quad (3.20)$$

avec  $it$  l'itération en cours.

### 5. Condition d'arrêt

Les exemplaires ne doivent pas changer pour les  $n$  dernières itérations de la boucle principale, où  $n$  est une valeur prédéfinie par l'utilisateur.

### 3.3. Analyse de l'influence des attributs et du choix de la métrique sur les résultats de la classification par AP

Les différentes métriques données dans le Tableau 6 ont été introduites pour calculer la matrice de similarité à l'entrée de l'AP et analyser leur influence sur le choix du paramètre de préférence et sur les résultats de partitionnement. Pour cette expérimentation, plusieurs images synthétiques ont été construites à partir de l'album de Brodatz [42]. Les Figures 10 et 11 montrent respectivement le procédé de construction de ces images et un échantillon de trois images test.

La figure 12 montre les résultats de partitionnement obtenus pour des images de la Figure 11. L'AP est utilisée ici dans sa version non supervisée. Le calcul des attributs est réalisé sur des fenêtres de taille 3x3 et la valeur du paramètre de préférence a été fixée à la valeur médiane de la matrice de similarité. Deux types d'attributs sont utilisés, les statistiques du premier ordre (4 attributs : moyenne, variance, skewness et kurtosis) puis les deux attributs issus des transformées en ondelettes (moyenne, énergie), les statistiques d'ordre deux (8 attributs). Cette figure montre également, le nombre de classes estimé automatiquement, ainsi que le taux moyen de bonne classification (ACCR en %) en fonction des attributs caractérisant les pixels et les indices de similarité utilisés.

Les Figures 13 et 14 donnent l'évolution du critère ACCR (%) en fonction des indices de similarité pour chacune des images, puis pour l'ensemble de ces images respectivement. Il est à noter que les meilleurs taux sont obtenus avec la distance L1 quelle que soit la nature des attributs utilisés et quelle que soit l'image à partitionner. Cette conclusion confirme les résultats de comparaison de différentes métriques donnés dans [147-148]. Cette métrique est donc retenue par la suite comme critère de similarité pour la formation des classes.

Tableau 6: Indices de similarité entre deux individus  $x_i$  et  $x_j$

Métriques	Formules
Distance L1 (où distance de Manhattan)	$S_{L1}(x_i, x_j) = -\ a_i - a_j\ _1$
Distance L2 (euclidienne)	$S_{L2}(x_i, x_j) = -\ a_i - a_j\ _2$
Distance $L_\infty$ (Tchebychev)	$S_{Che}(x_i, x_j) = -\max_{l \in \{1, Na\}}  a_{il} - a_{jl} $
Mesure de corrélation	$S_{Corr}(x_i, x_j) =  corr(a_i, a_j)  - 1$

Spectral Angle Mapper (SAM)	$S_{SAM}(x_i, x_j) = -\cos^{-1} \left( \frac{\langle a_i, a_j \rangle}{\ a_i\  \times \ a_j\ } \right)$
Mesure de divergence spectrale SID (Spectral Information Divergence)	$S_{SID}(x_i, x_j) = -\sum_{k=1}^{N_a} \sum_{l=1}^{N_a} \left( a'_{ik} \log\left(\frac{a'_{ik}}{a'_{jl}}\right) + a'_{jl} \log\left(\frac{a'_{jl}}{a'_{ik}}\right) \right)$ <p>avec <math>a'_i = \frac{a_i}{\ a_i\ }</math> et <math>a'_j = \frac{a_j}{\ a_j\ }</math></p>

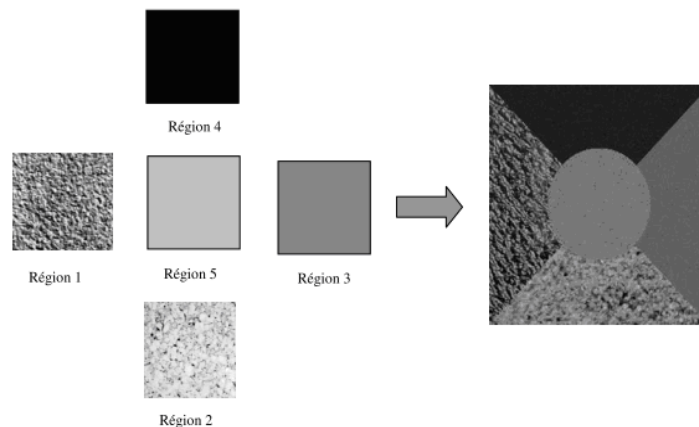


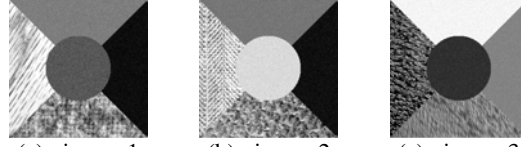
Figure 10: Procédé de construction d'une image synthétique

### 3.4. Discussion

L'état de l'art sur les différentes méthodes de classification présentées dans ce chapitre nous amène à formuler plusieurs remarques importantes :

- Les diverses méthodes classiques dites semi-supervisées (C-moyennes, C-moyennes floues, ISODATA) nécessitent toutes la connaissance *a priori* du nombre de classes en entrée, ce qui est difficile à obtenir dans le cas où l'on cherche à partitionner des images représentant des données naturelles, comme celles traitées en télédétection, pour lesquelles la diversité et la variabilité des classes sont importantes.
- La méthode de propagation d'affinité (AP) peut être utilisée en mode complètement non supervisé. Elle a été largement exploitée ces dernières années dans des domaines nombreux et divers. Les différentes études menées [114], [117], [118], [120] montrent sa supériorité quand elle est utilisée en mode semi-supervisé par rapport aux méthodes C-moyennes, C-moyennes floues et ISODATA. Elle présente cependant encore quelques limites liées à la dimensionnalité et à la surestimation du nombre des classes (sur-segmentation) quand elle est utilisée en mode non supervisé. Néanmoins grâce à

ses avantages et performances quant à la qualité des résultats de classification, mais aussi à son caractère déterministe, nous l'avons retenue comme méthode de classification. Dans la suite de cette thèse, nous apporterons des solutions pour remédier à ses inconvénients majeurs que sont la surestimation du nombre de classes et l'impossibilité de son application aux images hyperspectrales de grande taille spatiale.



(a) : image 1

(b) : image 2

(c) : image 3

Figure 11: Exemple d'images synthétiques à partitionner

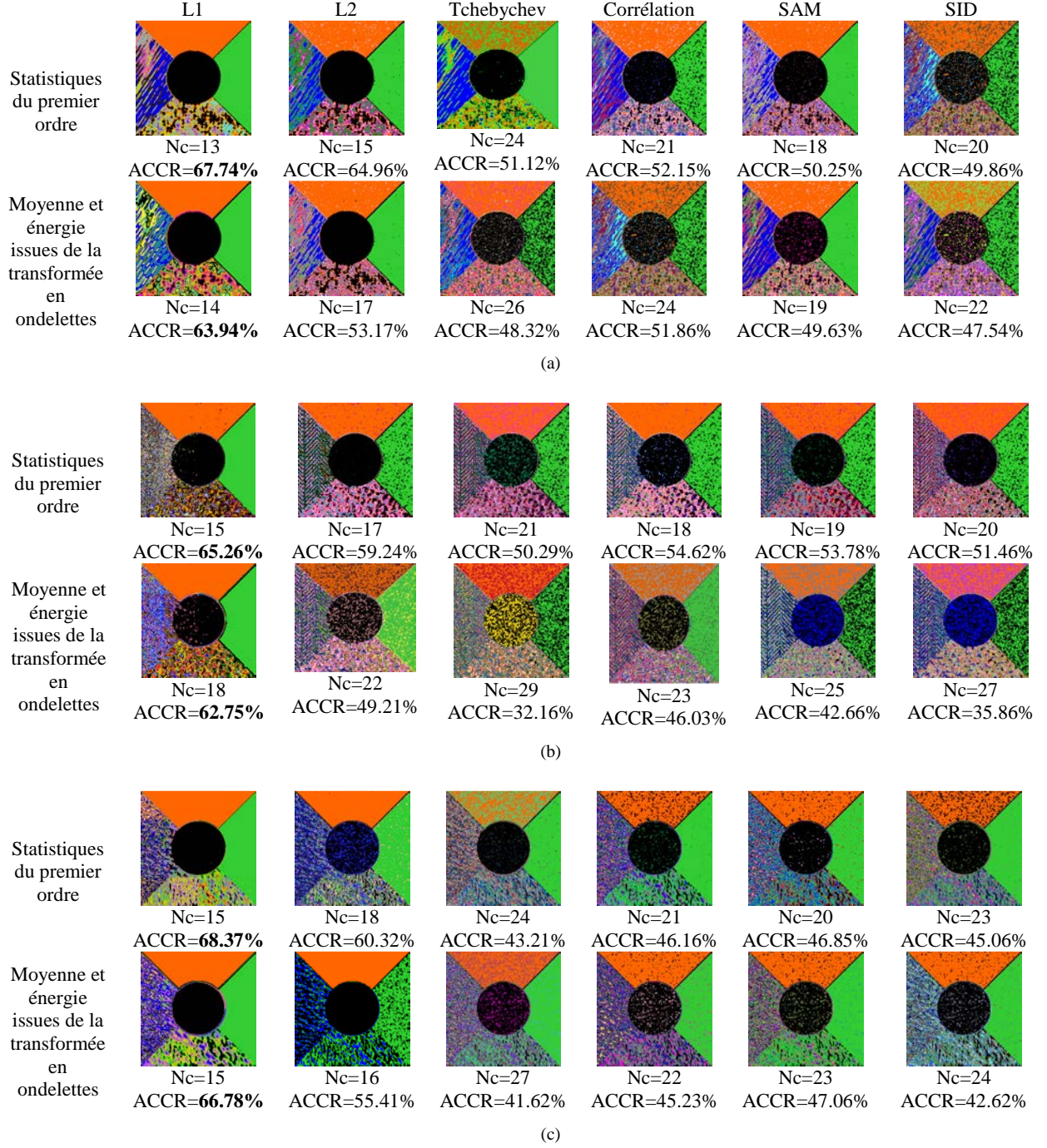
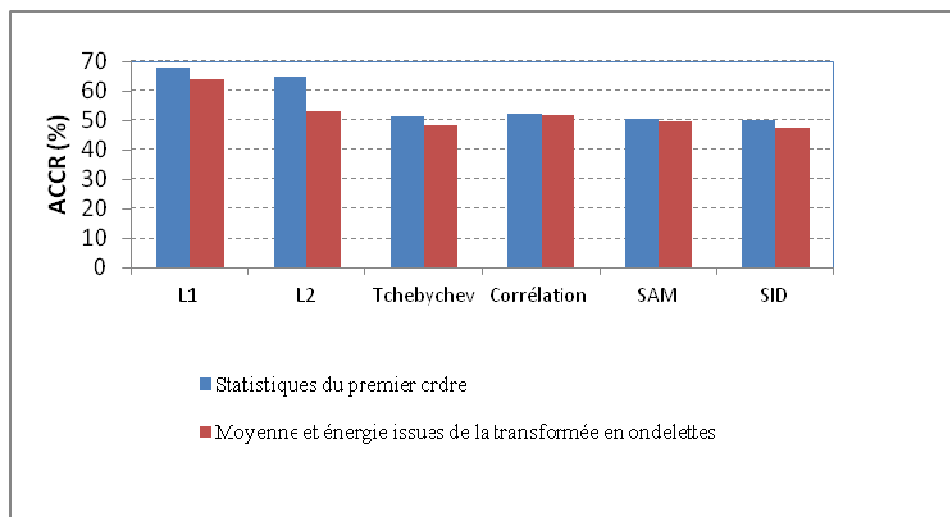
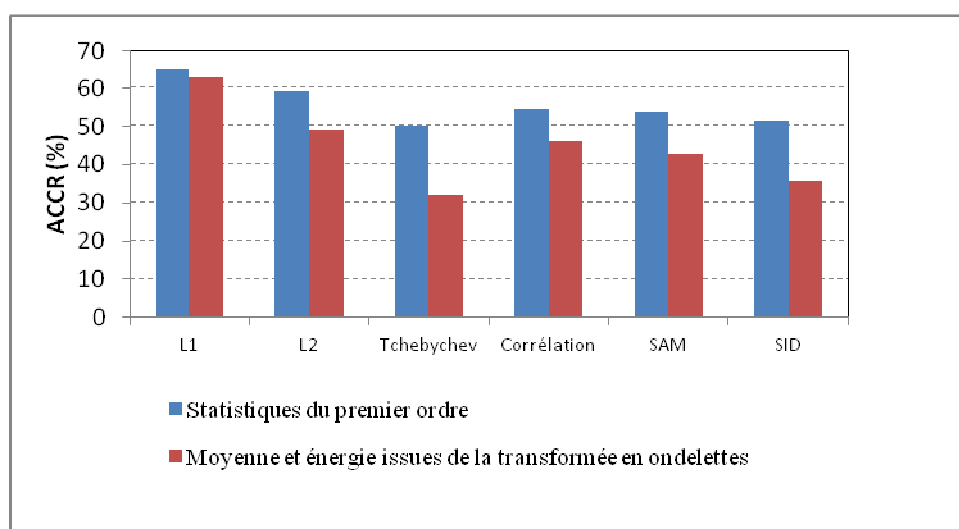


Figure 12: Résultats de partitionnement des images de la Figure 11 en fonction du choix de l'indice de similarité et des attributs.

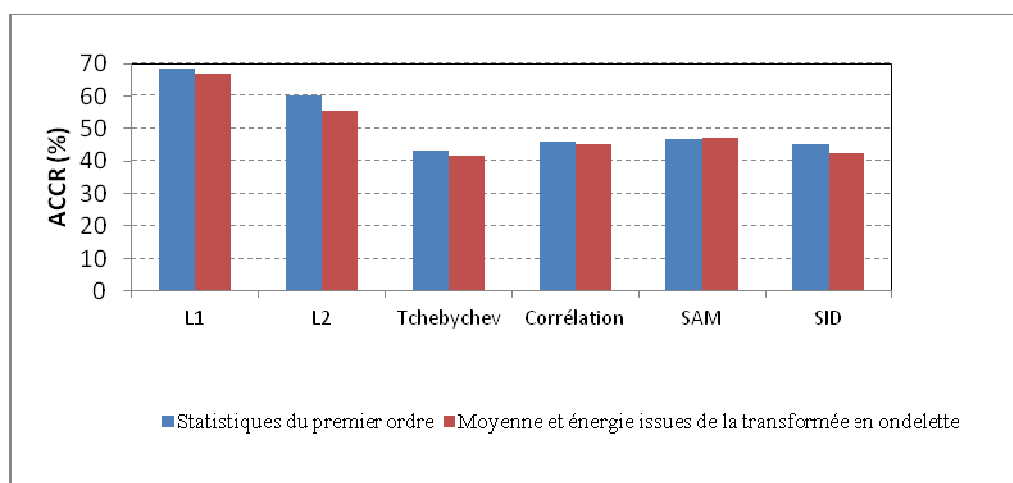




(a) : Résultat image 1



(b) : Résultat image 2



(c) : Résultat image 3

Figure 13: Taux de classification en fonction des indices de similarité

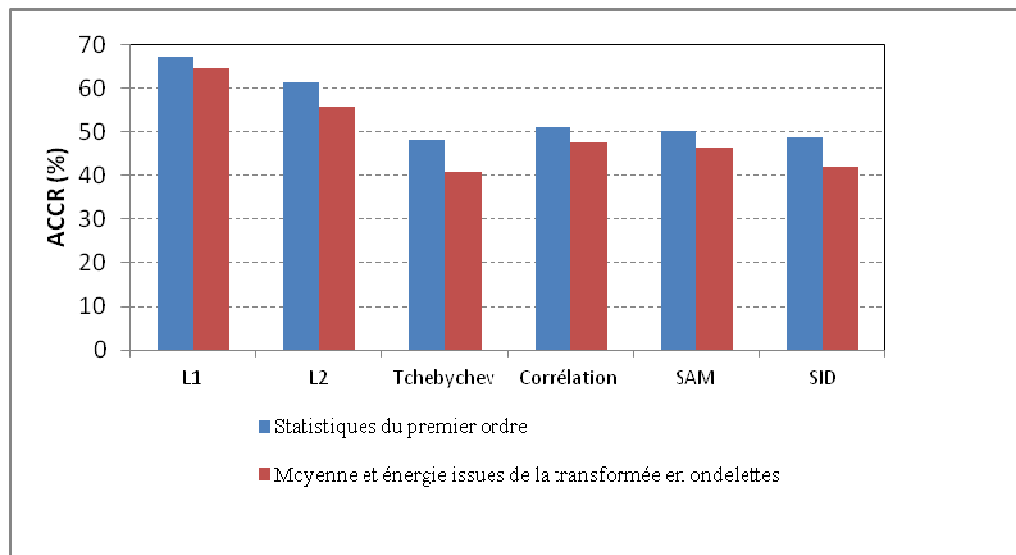


Figure 14: Taux de classification moyens pour les 3 images en fonction des indices de similarité

## Chapitre 4

# Critères d'évaluation d'une méthode de classification

### 4.1. Introduction

Les critères d'évaluation sont indispensables pour évaluer et quantifier les performances d'un algorithme de classification, car la qualité d'une partition conditionne fortement celle de l'interprétation. C'est pourquoi de nombreux critères sont proposés dans la littérature pour pouvoir évaluer la qualité des résultats de classification [133], [135], [137], [140], [144], [138]. Deux types d'évaluation peuvent être distingués :

- *Évaluation relative* : il s'agit d'évaluer la qualité d'un résultat de classification en mesurant sa similarité avec une vérité de terrain. Cette évaluation fait appel à des critères externes basés sur des connaissances *a priori*. Ce type d'évaluation pose le problème de la confiance dans la vérité de terrain, qui reste souvent subjective et approximative. Un exemple emblématique de ce problème est celui de l'image hyperspectrale AVIRIS Indian Pines, très largement utilisée dans la littérature en classification et dont la vérité de terrain associée contient un grand nombre d'erreurs et d'approximations patentées [146].

Dans le cas où une vérité de terrain fiable est disponible, la matrice de confusion peut être calculée.

- *Évaluation absolue* : des critères d'évaluation non supervisés sont utilisés pour estimer la qualité d'un résultat de partitionnement à partir de statistiques calculées sur chaque classe détectée. Ces critères dits « internes » sont basés sur les propriétés intrinsèques des données.

Dans ce chapitre, on se limite à l'analyse des principaux critères internes (non supervisés) d'évaluation des résultats de classification. Il est à noter que ces critères sont également utilisés comme indices d'agrégation pour affecter un individu ou pixel à une classe ou bien comme critère d'arrêt.

## 4.2. Critères d'évaluation non supervisés

L'étude réalisée par Voisine [40] pour la comparaison des méthodes de validation des algorithmes de classification permet de retenir les attributs suivants, considérés comme les plus pertinents : l'homogénéité (critère de Cochran), la compacité, et la stabilité. D'autres critères [45], [123], [124] ont montré souvent des résultats intéressants pour l'évaluation des résultats de classification. Nous rappelons ici ces critères.

- **Test d'homogénéité de Cochran**

Ce test donne une mesure de confiance sur l'homogénéité d'une région. Soit  $R$  une région formée d'un ensemble de sous-régions  $R_i$  telles que  $R = \bigcup_i R_i$

Le test d'homogénéité de Cochran [125] est défini par :

$$E_{cochran} = \frac{\max \sigma_i^2}{\sum_i \sigma_i^2} \quad (4.1)$$

Où  $\sigma_i^2$  est la variance de la région  $R_i$ . Cette mesure donne un critère fiable de l'homogénéité d'une région ou d'une classe. Elle permet de valider une région indépendamment des autres, et de plus, la variance peut être calculée sur n'importe quel ensemble d'attributs. Par contre la qualité de la validation est fortement dépendante du choix des attributs et du seuil.

- **La compacité**

La compacité est un paramètre morphologique global d'une région, qui mesure sa difformité. Ce paramètre est défini par :

$$E_{comp} = \frac{4\pi S(R)}{P^2(R)} \quad (4.2)$$

où  $S(R)$  et  $P(R)$  sont respectivement l'aire et le périmètre de la région  $R$ .

Cette mesure permet de bien détecter les régions compactes lorsqu'elles ne sont pas bruitées. Elle n'est toutefois pas efficace pour la validation de régions difformes ou allongées.

- **Critère adaptatif d'homogénéité ( $F_{CR}$ )**

Rosenberger et Chehdi [39], [41] ont proposé une méthode de validation basée sur deux critères adaptatifs. Ces deux critères reprennent les deux conditions de bonne classification de Haralick [44]. Le critère global d'évaluation défini (noté  $F_{CR}$ ) tient compte à la fois de l'homogénéité globale intra-classe et interclasse.

$$F_{CR} = \frac{\overline{D(I)} + 1 - D(I)}{2} \quad (4.3)$$

avec

$$\overline{D(I)} = \frac{1}{N_R} \sum_{i=1}^{N_R} \frac{r_i}{N} \overline{D(R_i)} \quad (4.4)$$

la disparité interclasse

et

$$\underline{D(I)} = \frac{1}{N_R} \sum_{i=1}^{N_R} \frac{r_i}{N} \underline{D(R_i)} \quad (4.5)$$

la disparité intra-classe

où  $N$ ,  $N_R$ ,  $R_i$  et  $r_i$  représentent respectivement le nombre de pixels, le nombre de régions dans l'image, une région  $i$  et le nombre de pixels dans la région  $R_i$ .

Les deux mesures  $\underline{D(R_i)}$  et  $\overline{D(R_i)}$  dépendent du type de régions segmentées. Si une région est considérée comme uniforme, ces mesures dépendent de l'écart type de ses niveaux de gris. Si elles sont texturées, elles dépendent d'un ensemble d'attributs de texture. Cette mesure donne une évaluation globale de la segmentation, par contre elle ne permet pas de comparer la qualité de deux régions.

- **Critère d'évaluation de Levine et Nazif**

Levine et Nazif [140] proposent une mesure de performance de la classification qui peut être calculée en temps réel. Ce critère calcule la somme des contrastes des régions  $R_i$  pondérées par leurs aires  $S(R_i)$ . Le contraste d'une région est défini à partir des contrastes existants avec les régions adjacentes.

$$Inter-LN = \frac{\sum_{R_i} S(R_i) LN_i}{\sum_{R_i} S(R_i)} \quad (4.6)$$

$$\text{avec } LN_i = \sum_{R_j} \frac{l_{ij}}{P(R_i)} \left| \frac{\mu_i - \mu_j}{\mu_i + \mu_j} \right| \quad (4.7)$$

Avec  $\mu_i$  la moyenne dans la région  $R_i$  ;  $l_{ij}$  la longueur de la frontière commune entre  $R_i$  et  $R_j$  ; et  $P(R_i)$  le périmètre de la région  $R_i$ .

Plus cette mesure est élevée, meilleure est l'évaluation de la qualité des résultats de classification.

- **Contraste de Zeboudj**

Zeboudj [126], [127] définit une mesure de contraste d'une région dans l'image, qui reflète d'une part l'homogénéité intérieure de la région, et d'autre part, l'importance des transitions avec les régions voisines. Cette mesure est utilisée pour sélectionner les meilleurs seuils dans un processus de classification. Si on note  $V_x$  un voisinage d'un pixel  $x$ , le contraste intérieur et extérieur d'une région  $R_i$  d'aire  $S(R_i)$  et de frontières  $F_i$  se définissent respectivement par :

$$CI_i = \frac{1}{S(R_i)} \sum_{x \in R_i} \max(d(x, y), y \in V_x \cap R_i) \quad (4.8)$$

$$CE_i = \frac{1}{l_i} \sum_{x \in R_i} \max(d(x, y), y \in V_x, y \notin R_i) \quad (4.9)$$

où  $l_i$  est la longueur de  $F_i$ .

Le contraste de  $R_i$  est donné par :

$$C(R_i) = \begin{cases} 1 - \frac{CI_i}{CE_i}, & \text{si } 0 < CI_i < CE_i \\ CE_i & \text{si } CI_i = 0 \\ 0 & \text{sinon} \end{cases} \quad (4.10)$$

Le contraste global est finalement :

$$C_{Zeboudj} = \frac{1}{S} \sum_i S(R_i) C(R_i) \quad (4.11)$$

avec  $S$  l'aire totale des régions.

- **Statistique de Huberts**

Dans [128], l'indice  $\Gamma$  (aussi appelé “ statistique de Huberts”) est proposé pour évaluer un algorithme de classification. Considérons l'ensemble  $X$  des individus à classer et  $N' = N(N-1)/2$  le nombre de paires possibles dans  $X$ . Cet indice est défini par:

$$\Gamma = \frac{1}{N'} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_1(x_i, x_j) \times d_2(x_i, x_j) \quad (4.12)$$

Trois cheminements sont alors envisagés pour évaluer ou comparer entre eux des schémas de classification.

- **Évaluation Externe** : dans ce cas  $d_1(x_i, x_j)$  et  $d_2(x_i, x_j)$  correspondent aux distances entre les classes contenant les individus  $x_i$  et  $x_j$ , respectivement dans le schéma obtenu et dans la classification issue des connaissances externes (classification prédéfinie). L'optimum est atteint pour deux schémas identiques, caractérisés par :

$$\Gamma = \frac{1}{N'} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_2(x_i, x_j)^2 \quad (4.13)$$

- **Évaluation Interne** : cet indice permet d'évaluer l'adéquation entre un résultat de classification  $Cl$  et la matrice de dissimilarité  $Diss$  dont il est issu. Pour cela on pose  $d_1(x_i, x_j) = diss(x_i, x_j)$  (dissimilarité entre les deux individus) et  $d_2(x_i, x_j) = 1$  si  $x_i$  et  $x_j$  appartiennent à des classes différentes dans  $Cl$ , et 0 sinon.

Le maximum est atteint pour le résultat  $Cl = \{\{x_1\}, \dots, \{x_N\}\}$  (chaque classe est constituée d'un unique individu). En revanche, pour un nombre fixe de classes, l'optimum (maximum) caractérise un résultat constitué de classes bien séparées relativement à la matrice  $Diss$ .

- **Évaluation Relative** : dans ce dernier cas  $d_1(x_i, x_j)$  correspond toujours à la distance entre les deux objets  $x_i$  et  $x_j$ , tandis que  $d_2(x_i, x_j)$  mesure la distance entre les classes contenant  $x_i$  et  $x_j$ . Pour deux objets appartenant à une même classe, le produit est nul ( $d_2(x_i, x_j) = 0$ ). On

cherche alors le résultat de classification qui maximise ce critère, révélant ainsi des classes compactes et séparées.

Un grand nombre d'autres critères mesurant en général le degré d'homogénéité ou d'hétérogénéité des classes sont aussi couramment utilisés:

- Intra [129]: ce critère indique le degré d'homogénéité entre les individus  $x$  et  $y$  appartenant à la même classe  $C_i$ . Ce critère est à minimiser.

$$Intra = \frac{1}{N_c} \sum_{C_i} \frac{1}{2N_{C_i}} \sum_x \sum_y d^2(x_i, x_j) \quad (4.1 \ 4)$$

- Inter[129] : ce critère mesure le degré d'hétérogénéité ou de séparabilité entre les classes. Il est à maximiser.

$$Inter = \frac{1}{N_c} \sum_{C_i} N_{C_i} d^2(g_i, g_x)^2 \quad (4.1 \ 5)$$

où  $g_x$  est le centre de gravité de l'ensemble des pixels.

- Root-mean-square standard deviation [130]: ce critère permet également la mesure de l'homogénéité des classes formées. Il est à minimiser.

$$RMSSTD = \left\{ \sum_i \sum_{x \in C_i} \|a_x - g_i\|^2 / [N_a \sum_i (N_{C_i} - 1)] \right\}^{\frac{1}{2}} \quad (4.1 \ 6)$$

avec  $a_x$  le vecteur d'attributs représentant le pixel  $x$ .

- R-squared [130] : il mesure le degré de différence entre les classes.

$$RS = (\sum_{x \in X} \|a_x - g_X\|^2 - \sum_i \sum_{x \in C_i} \|a_x - g_i\|^2) / \sum_{x \in X} \|a_x - g_X\|^2 \quad (4.1 \ 7)$$

avec  $g_x$  est le centre de gravité de  $X$ , et  $g_i$  le centre de gravité de  $C_i$ .



- Calinski-Harabasz index [131] : ce critère pondère la variance intra-classe par le nombre de classes et permet donc d'avoir un nombre "optimal" de classes de manière objective. Sa maximisation correspond à la partition optimale.

$$CH = \frac{\sum_i N_{C_i} d^2(g_i, g_x) / (Nc - 1)}{\sum_i \sum_{x \in C_i} d^2(a_x, g_i) / (N - Nc)} \quad (4.18)$$

- Indice de Davies-Bouldin [132] : il est basé sur la minimisation du rapport des dispersions intra-classe et de la séparation interclasse. La meilleure partition est celle qui minimise la moyenne de la valeur calculée pour chaque classe. En d'autres termes, la meilleure partition est celle qui minimise la similarité entre les classes.

$$DB = \frac{1}{Nc} \max_{\substack{1 \leq j \leq Nc \\ j \neq i}} \left\{ \frac{S_a(C_i) + S_a(C_j)}{d_a(C_i, C_j)} \right\} \quad (4.19)$$

avec  $S_a$  la séparation interclasse et  $d_a$  la dispersion intra-classe

- Indice de Xie-Beni [133]: il définit la séparation interclasse comme le minimum du carré de la distance entre les centres des classes et la compacité intra-classe comme la moyenne du carré de la distance entre l'ensemble des données et les centres des classes correspondants. Le nombre optimal de classes (partition optimale) est celui qui minimise ce critère.

$$XIE = \sum_i \sum_{x \in C_i} d^2(a_x, g_i) / [N \times \min_{i, j \neq i} d^2(g_i, g_j)] \quad (4.20)$$

- Indice de SD validity [134]: ce critère est basé sur le concept de dispersion moyenne et de séparation totale entre les classes. Le premier terme évalue la compacité basée sur la variance des classes et le second terme évalue la différence de séparation basée sur la distance entre les centres des classes. La valeur de cet indice est la somme de ces deux termes, le nombre optimal de classes peut être obtenu en minimisant la valeur de SD.

$$\begin{aligned}
SD &= Scat(N_c) + Dis(N_c) \\
Scat(N_c) &= \frac{1}{N_c} \sum_i \|\sigma_{C_i}\| / \|\sigma_X\| \\
Dis(N_c) &= \frac{\max_{i,j} d(g_i, g_j)}{\min_{i,j} d(g_i, g_j)} \sum_i (\sum_j d(g_i, g_j))^{-1}
\end{aligned} \tag{4.21}$$

avec  $\sigma_{C_i}$  la matrice de covariance de  $C_i$ , et  $\sigma_X$  celle de  $X$

- Indice de validité  $S\_Dbw$  [135] : cet indice prend en compte la densité pour mesurer la séparation interclasse. L'idée de base est que pour chaque paire de centres de classe, au moins une de leurs densités doit être plus grande que la densité de leur point médian. La compacité intra-classe est la même que dans  $SD$ . De même, l'indice est la somme de ces deux termes et la valeur minimale de  $S\_Dbw$  indique le nombre optimal des classes.

$$\begin{aligned}
SDbw &= Scat(Nc) + Dsbw(Nc) \\
Dsbw(Nc) &= \frac{1}{Nc} \sum_i \left[ \sum_{j, j \neq i} \frac{\sum_{x \in C_i \cup C_j} f(a_x, u_{ij})}{\max \left\{ \sum_{x \in C_i} f(a_x, c_i), \sum_{x \in C_j} f(a_x, c_j) \right\}} \right]
\end{aligned} \tag{4.22}$$

où  $u_{ij}$  est la médiane de  $g_i$  et  $g_j$  les centres de gravité des classes. La fonction de densité autour de  $g_i$  se définit comme suit : elle compte le nombre d'individus dans un hyper-sphère dont le rayon est égal à la moyenne des écart-type des classes. L'écart type moyen des classes est défini par :

$$stdev = \frac{1}{Nc} \sqrt{\sum_{i=1}^{Nc} \|\sigma(g_i)\|} \tag{4.23}$$

- Indice de Silhouette [136]: cet indice procède à l'échelle microscopique, c'est à dire qu'il s'intéresse aux individus en particulier et non pas aux classes. Le but de ce critère est de vérifier si chaque individu a été bien classé.

$$\begin{aligned}
Sil &= \frac{1}{N_c} \sum_i \{N_{C_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max[b(x), a(x)]}\} \\
\text{avec } a(x) &= \frac{1}{N_{C_i}} \sum_{y \in C_i, y \neq x} d(x, y) \text{ et } b(x) = \min_{j, j \neq i} \left[ \frac{1}{N_{C_j}} \sum_{y \in C_j} d(x, y) \right]
\end{aligned} \tag{4.24}$$

- **Indice de Dunn [137]** : Ce critère est à maximiser en calculant le rapport entre la distance maximum entre deux individus classés ensemble et la distance minimum qui sépare deux éléments classés différemment.

$$Dunn = \min_i \left\{ \min_j \left( \frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \left\{ \max_{x, y \in C_k} d(x, y) \right\}} \right) \right\} \quad (4.25)$$

Ces critères sont dans la plupart des cas efficaces et simples à utiliser pour évaluer un résultat de classification.

Les premiers critères cités ont été utilisés dans les thèses de Rosenberger [39] et Moghrani [138]. Certains de ces critères ont également été utilisés dans [123] comme le critère de Levine et Nazif et le critère de Yasnoff. D'autres critères ont été définis dans la thèse de Voisine [40] comme la stabilité et le temps de calcul.

### 4.3. Analyse des indices d'évaluation associés à l'AP.

Dans cette section, une analyse de neuf critères parmi ceux cités dans la section précédente et précisés dans le tableau de l'annexe B est menée pour examiner leur comportement en fonction de l'évolution du paramètre de préférence de la méthode AP, paramètre vis-à-vis duquel elle est très sensible. Pour cette analyse, nous avons utilisé des données issues d'une image aérienne réelle hyperspectrale de taille 8075x9748 pixels, en 62 bandes spectrales couvrant le spectre visible/proche infrarouge [400, 970] nm. La résolution spatiale au sol de cette image est de 0.50m. Cette image a été acquise dans la région de Murcia (Espagne) en octobre 2010 à l'aide du spectromètre imageur AISA Eagle de l'équipe TSI2M dans le cadre d'un projet en partenariat avec INFRAECO, entreprise hispano-chilienne spécialisée dans la gestion environnementale. L'image hyperspectrale que nous avons traitée est de petite dimension spatiale (64x64). Elle a été construite à partir d'un ensemble de régions pour lesquelles nous disposons d'une vérité de terrain, c'est-à-dire mesures spectrales et observations. Nous avons sélectionné cinq classes parmi celles de la vérité de terrain pour construire cette image. À cette image, nous avons associé une vérité de terrain comme le montre la Figure 15. Les données de l'image (c) ont été prélevées aléatoirement à partir des zones 1, 2, 3, 4 et 5. Ces zones correspondent respectivement aux classes « Rivière », « *Pinus halepensis* », « Pêchers », « *Arundo donax* » et « Bâtiment ».

Comme le montre la Figure 16, les signatures spectrales des pixels au sein des différentes classes présentent des chevauchements importants et nous pouvons remarquer que certaines classes ne sont pas homogènes physiquement. Cette hétérogénéité réelle est susceptible d'influencer fortement le taux de classification global.

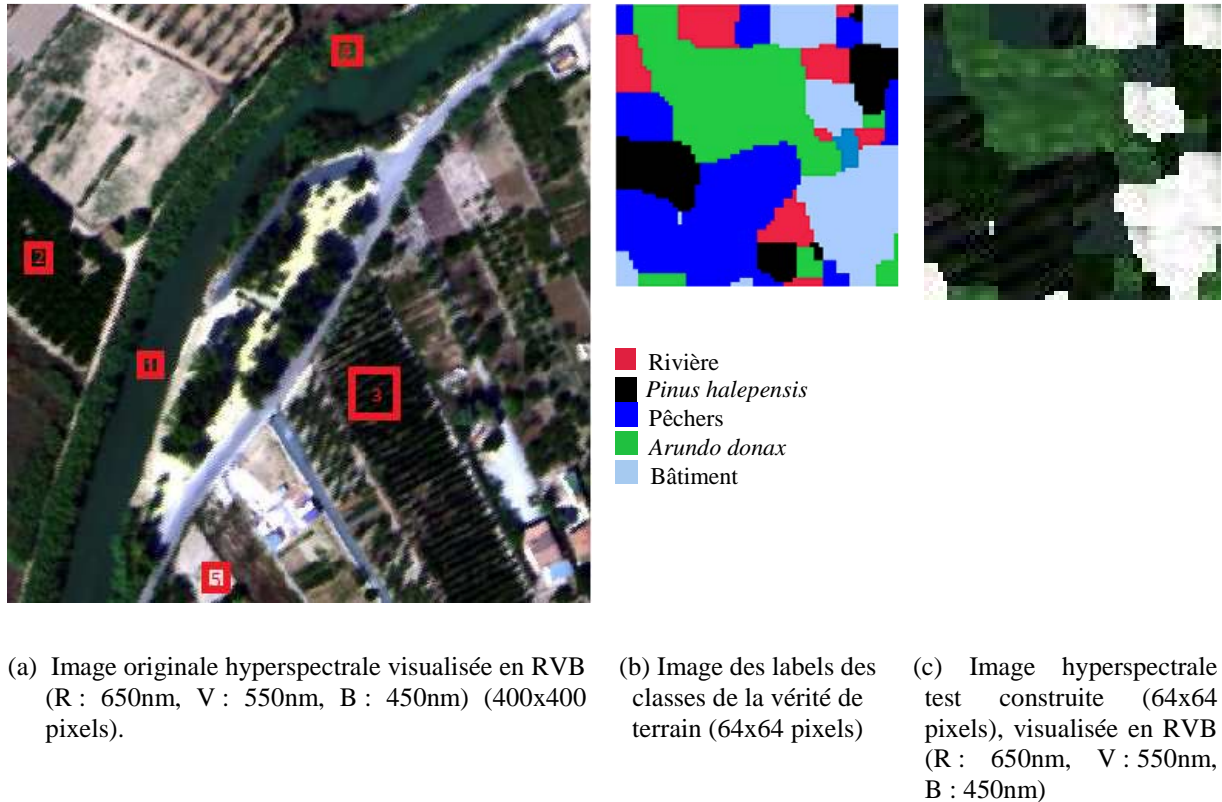


Figure 15: Image originale et image test construite

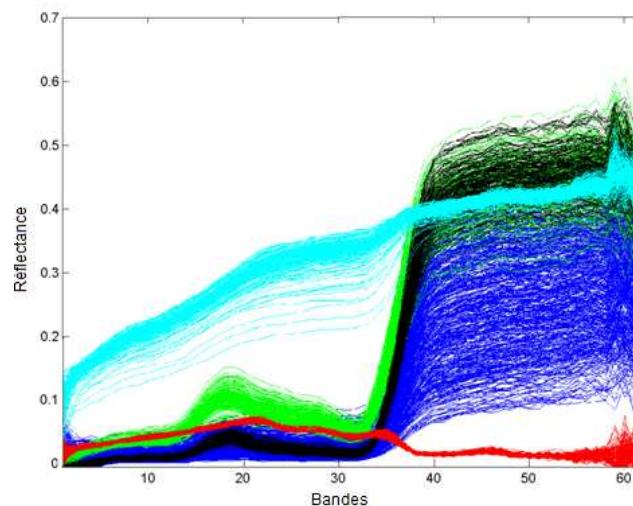


Figure 16: Les signatures spectrales de la vérité de terrain des cinq classes de l'image hyperspectrale construite.

Pour cette analyse, nous avons mesuré l'évolution des neuf indices d'évaluation de la qualité d'une partition en fonction des attributs et du paramètre de préférence  $p$ . Les Figures 17, 18 et 19 représentent ces évolutions pour les différents jeux d'attributs évoqués précédemment. L'analyse de ces évolutions est résumée dans le Tableau 7 qui donne la partition optimale en fonction du critère choisi. Ces résultats montrent que le critère de Levine et Nazif (Inter-LN) donne les résultats les plus probants à la fois pour le taux moyen de bonne classification et pour l'estimation du nombre exact de classes présentes dans les données. Ce critère sera par conséquent retenu dans le développement de l'approche de classification présentée dans la seconde partie de ce mémoire.

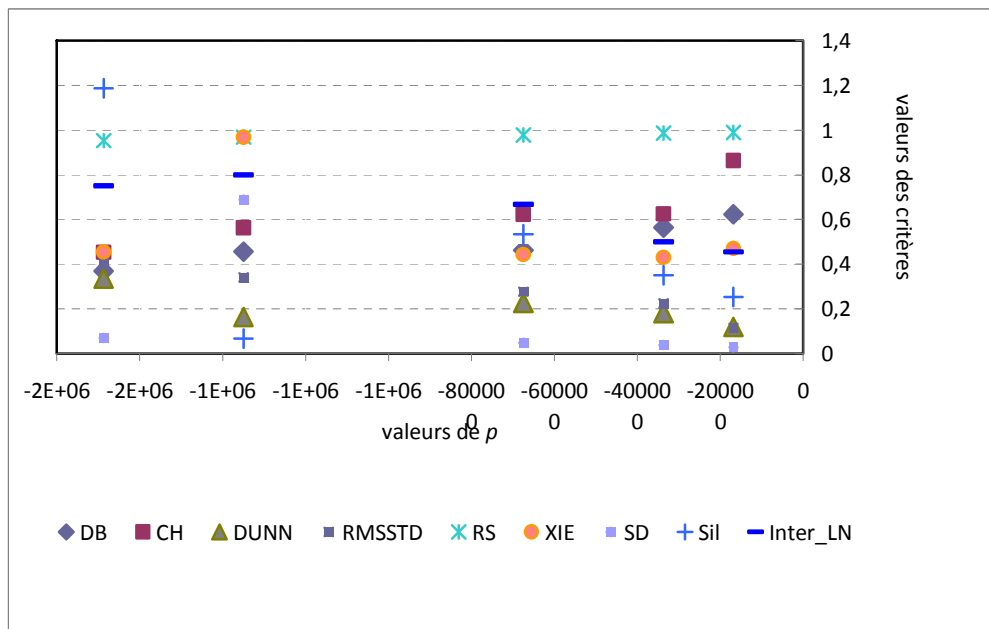


Figure 17: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la Figure 15(c) (attribut : signature spectrale d'un pixel)

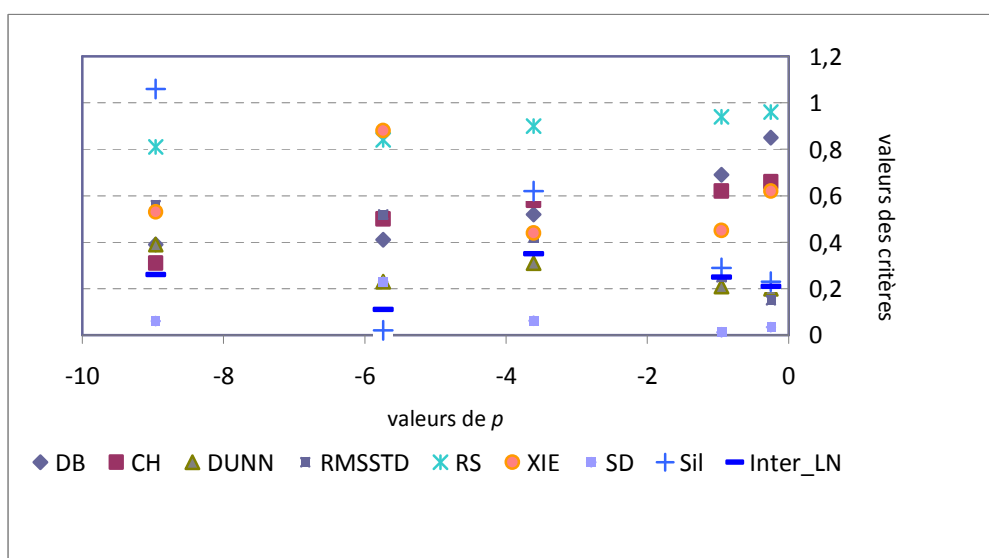


Figure 18: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la Figure 15(c) (attributs : statistiques du 1<sup>er</sup> ordre calculés sur une fenêtre 3x3)

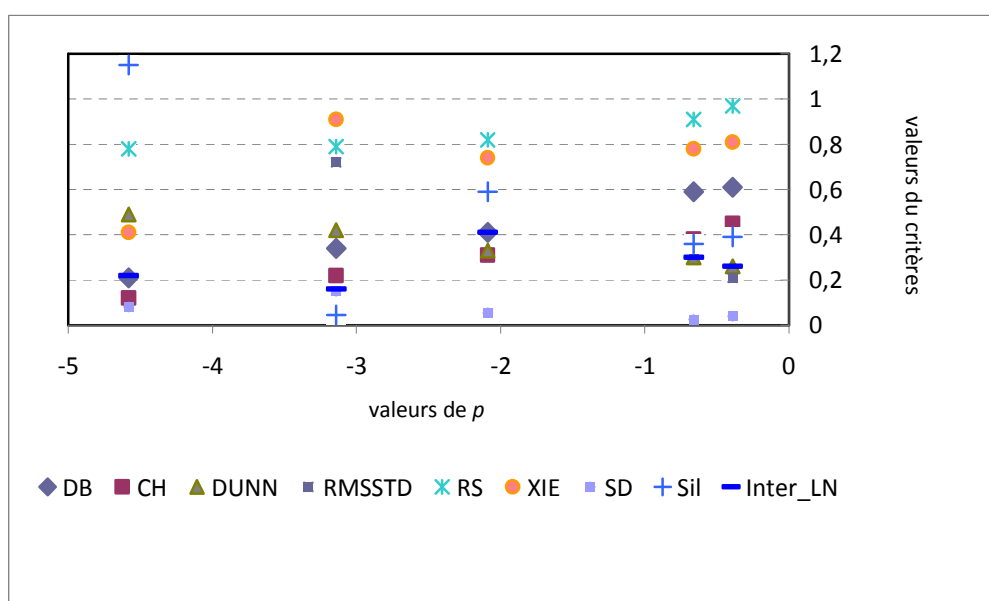


Figure 19: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la figure 15(c) (attributs : moyenne et énergie des coefficients d'ondelettes calculés sur une fenêtre 3x3).

Tableau 7: Valeur optimale du paramètre de préférence, nombre de classes obtenues et ACCR correspondant pour différents jeux d'attributs et différents critères d'évaluation.

<i>Indice</i>		Signature spectrale	Statistiques du premier ordre	Moyenne et énergie issues de la transformée en ondelettes
<b>Inter_LN</b>	<i>p</i> optimal	-1350 10 <sup>3</sup>	-3,612	-2.09
	Nc	<b>5</b>	<b>5</b>	<b>5</b>
	ACCR (%)	<b>90.72</b>	<b>91.94</b>	<b>89.65</b>
SIL	<i>p</i> optimal	-1687 10 <sup>3</sup>	-8,96	-4,58
	Nc	4	3	3
	ACCR (%)	71.69	52.49	47.79
SD	<i>p</i> optimal	-1687 10 <sup>3</sup>	-0,95	-0,66
	Nc	11	8	9
	ACCR (%)	49.17	62.04	53.17
XIE	<i>p</i> optimal	-337 10 <sup>3</sup>	-3,612	-4,58
	Nc	8	<b>5</b>	3
	ACCR (%)	60.33	<b>91.94</b>	47.79
RS	<i>p</i> optimal	-1687 10 <sup>3</sup>	-8,96	-4,58
	Nc	4	3	3
	ACCR (%)	71.69	52.49	47.79
RMSSTD	<i>p</i> optimal	-16810 <sup>3</sup>	-0,25	-0,39
	Nc	11	9	11
	ACCR (%)	49.17	51.29	45.48
DUNN	<i>p</i> optimal	-168 10 <sup>3</sup>	-0,25	-0,39
	Nc	11	9	11
	ACCR (%)	49.17	51.29	45.48
CH	<i>p</i> optimal	-168 10 <sup>3</sup>	-0,25	-0,39
	Nc	11	9	11
	ACCR (%)	49.17	51.29	45.48
DB	<i>p</i> optimal	-1687 10 <sup>3</sup>	-8,96	-4,58
	Nc	4	3	3
	ACCR (%)	71.69	52.49	47.79

## Conclusion de la Partie I

L'état de l'art présenté dans cette partie nous a permis d'appréhender les avantages et les inconvénients de chacune des méthodes évoquées de caractérisation, de classification des pixels, deux étapes centrales dans la chaîne d'analyse et d'interprétation de l'information, ainsi que d'analyser les performances comparées de techniques non supervisées d'évaluation de la classification. Cette analyse nous a également permis de proposer et de justifier les grandes lignes de l'approche qui va être proposée dans ce mémoire.

Cette analyse nous a conduit aux conclusions suivantes :

- **Concernant l'étape de caractérisation des pixels des images hyperspectrales :**
  - les attributs classiques pour la caractérisation des pixels des images monocomposantes peuvent être adaptés aux images hyperspectrales en les calculant bande par bande.
  - les statistiques d'ordre supérieur permettent une meilleure caractérisation des pixels des images. Cependant, nous constatons que cette étape est réalisée, dans la majorité des cas, sans réellement se soucier de l'information spectrale. Or, il nous semble que dans une démarche de partitionnement d'image hyperspectrale, l'exploitation conjointe de l'information spatiale et spectrale apparaît tout à fait légitime voire même indispensable. En effet, les approches qui combinent l'information spatiale et spectrale donnent des résultats de classification meilleurs que les approches qui ne tiennent compte que de l'information spectrale.
- **Concernant l'étape de classification :** plusieurs algorithmes de classification d'images hyperspectrales ont été proposés dans la littérature. La comparaison des uns par rapport aux autres est une tâche extrêmement difficile, car dans la plupart des cas les résultats engendrés sont variables en fonction des applications considérées. Une autre difficulté provient du fait qu'il n'existe pas de critères rigoureux de sélection d'un algorithme particulier. Les conclusions que l'on peut dégager des différentes analyses de cet état de l'art sont les suivantes:



- Les approches utilisant les notions d'analyse de données paraissent les plus robustes pour le traitement des données de grandes dimensions.
- les diverses méthodes de classification semi-supervisées (C-moyennes, C-moyennes floues, ISODATA) ne sont pas appropriées pour le partitionnement des données réelles, puisque toutes nécessitent a minima la connaissance *a priori* du nombre de classes.
- La méthode de classification non supervisée par propagation d'affinité a été largement exploitée ces dernières années dans des domaines nombreux et divers. En revanche, cette méthode dans sa version de base présente encore quelques limites liées à la dimensionnalité en termes de nombre d'individus et au sur-partitionnement de l'image. Pour remédier à ces problèmes, une approche non supervisée de classification des données de grande dimension sera présentée dans le chapitre suivant.
- **Enfin, concernant l'étape d'évaluation des méthodes de classification**, celle-ci est indispensable pour quantifier les performances des algorithmes. Plusieurs méthodes d'évaluation ont été présentées dans la littérature et qui peuvent être utilisées, on peut citer les plus pertinents et les plus utilisés ; la compacité, la stabilité des résultats, et les critères d'homogénéité des classes d'une partition. Nous avons retenu le critère de Levine et Nazif après une analyse expérimentale comparative des performances de différents critères menée dans la section 4.2. Il sera introduit comme critère d'optimisation dans la procédure d'estimation du nombre de classes issues de l'application de la méthode de classification AP.

## **Partie II**

# **Approche de partitionnement développée et évaluation**

Cette partie est structurée en deux chapitres. Dans le premier (chapitre 5), nous présentons l'approche proposée de caractérisation et de classification des pixels. Tout d'abord, nous introduisons et analysons les différents attributs développés pour des images multi-hyperspectrales. Nous présentons ensuite la méthode de réduction des données à classer et la méthode de propagation par affinité (AP) optimisée. Le chapitre 6 porte sur l'évaluation et la validation de la chaîne d'analyse et de partitionnement des images multi et hyperspectrales développée sur trois applications réelles : identification de plantes invasives et identification du couvert paysager, dont la détection des cèdres, à partir d'images aéroportée et satellitaire.

# Chapitre 5 : Approche proposée

## 5.1. Introduction

L'état de l'art présenté dans le premier chapitre nous a permis d'évoquer les particularités et les atouts des méthodes de caractérisation et de classification des pixels. Pour l'extraction des attributs, nous avons souligné l'intérêt des statistiques d'ordre supérieur pour la caractérisation des pixels des images complexes. Quant à la classification des pixels, les méthodes de classification non supervisées et non paramétriques présentent un intérêt pour l'utilisateur car les connaissances du terrain ne sont pas exigées. La méthode de propagation d'affinité répond parfaitement à cet objectif.

Tenant compte de ces considérations, nous présentons dans ce chapitre l'architecture générale de l'approche proposée en décrivant les deux modules qui la composent (cf. Figure : 20)

- *Le premier module* est l'extraction d'attributs statistiques pour la caractérisation des pixels des images monocomposantes ou hyperspectrales. Une nouvelle approche de calcul d'attributs de tri-occurrences calculés sur des voisinages cubiques est proposée. Ces attributs tiennent compte de l'aspect spatial et spectral des données. Une étude comparative est menée afin de tester le pouvoir discriminant de ces attributs par rapport à d'autres attributs de la littérature.
- *Le second module* est la classification des pixels de l'image. Deux étapes sont introduites pour pouvoir appliquer l'approche de classification par affinité à des données de grande dimension :
  - Réduction du nombre d'individus à classer pour permettre l'exécution de l'AP.
  - Optimisation et application de l'AP sur les données retenues.

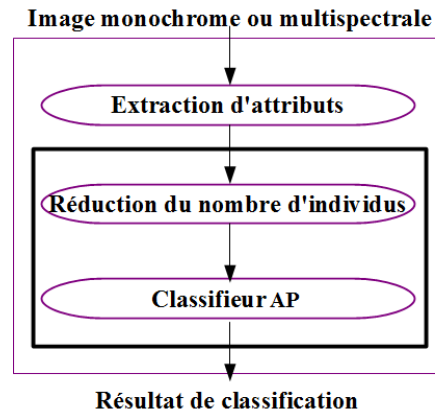


Figure 20: Chaîne de partitionnement d'images

Nous détaillons dans les sections suivantes les différentes étapes de chaque module et nous menons une étude comparative sur des images réelles et synthétiques pour évaluer l'approche proposée.

## 5.2. Extraction d'attributs de tri-occurrences

Pour caractériser un pixel, les attributs statistiques sont souvent estimés en utilisant les statistiques du premier ou du second ordre. Les différentes analyses menées dans l'état de l'art montrent que l'introduction des statistiques d'ordre supérieur contribue à la discrimination des classes de textures complexes. Par conséquent, nous introduisons les statistiques d'ordre trois.

Le calcul des attributs de tri-occurrences est basé sur l'extension du calcul des attributs de cooccurrences à l'ordre deux. Le schéma de calcul de ces attributs suit donc les mêmes étapes que celles des cooccurrences, à savoir :

- Réduction du nombre des niveaux, bande par bande par la fonction de multi-seuillage proposée par Kermad et Chehdi [139],
- Choix du voisinage  $V_{xi}$  suivant le type d'image : voisinage 2D ou 3D,
- Calcul des matrices de tri-occurrences : matrice 3D des fréquences d'apparition de triplets de valeurs de pixels organisés de manière fixe dans le voisinage  $V_{xi}$ ,
- Calcul des attributs d'Haralick : contraste, corrélation, moment des différences inverses, moyennes des sommes, variance des sommes, entropie, nuance du groupe (cluster shade) et proéminence du groupe (cluster prominence).

Pour le calcul de ces attributs, nous avons proposé deux schémas : le premier est adapté au cas des images monochromes ou monobandes, et le second au cas des images multi-hyperspectrales. La méthode de tri-occurrences sera aussi comparée avec différentes méthodes de calcul d'attributs : signatures spectrales, moments statistiques (moyenne, variance, skewness et kurtosis), la moyenne et l'énergie calculées à partir de la transformée en ondelettes, et enfin les attributs classiques de la cooccurrence.

La validation de l'étape de calcul d'attributs passe impérativement par la classification pour tester le pouvoir discriminant des paramètres calculés. Nous utilisons dans cette section la version standard de l'AP non supervisée ; les modifications réalisées seront présentées dans la section suivante. Pour pouvoir exécuter la version standard de l'AP, l'étape de validation de la pertinence des attributs ne portera que sur des images de petites tailles.

#### 5.2.1. Attributs de tri-occurrences spatiaux

Inspirés par l'approche proposée dans [61] pour la caractérisation des pixels des images texturées, basée sur les statistiques d'ordre trois, nous proposons de l'étendre au calcul des attributs d'Haralick. Dans [61], Reuzé propose de caractériser les pixels par des attributs (somme des contrastes, produit des contrastes, homogénéité du contraste) calculés à partir d'un couple de distances  $(d_1, d_2)$ . Chaque couple  $(d_1, d_2)$  définit les distances entre le point central d'une fenêtre de taille 3x3 et deux autres pixels localisés différemment dans cette même fenêtre. On distingue alors 28 combinaisons possibles de disposition de triplets, constituées de 5 formes orientées dans 4 directions (cf. figure 21). Le calcul des trois attributs précédents sur quatre images de l'album de Brodatz ayant les micro-textures les plus difficiles à discriminer et l'application du critère de minimisation de la distance de Mahalanobis pour la classification donne un taux de bonne classification de 100%.

En considérant le même schéma de disposition des trois pixels dans l'image présentée par la Figure 21, nous proposons de calculer la fréquence d'apparition des trois niveaux sur des fenêtres de taille (3x3). Nous nous sommes limités à des fenêtres de petite taille afin d'éviter une dégradation du taux de classification.

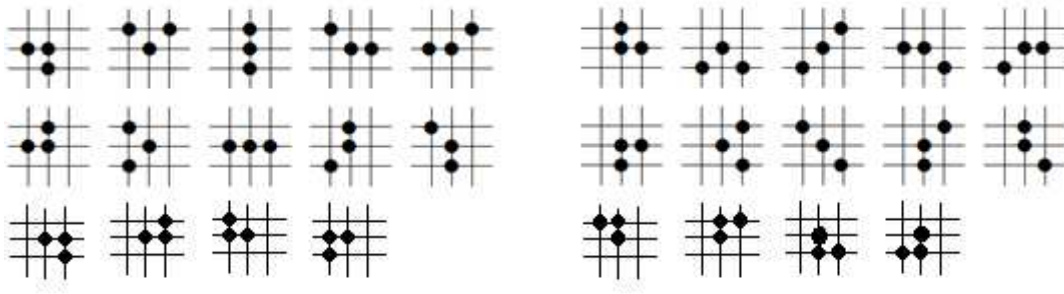


Figure 21: Les 28 combinaisons spatiales possibles

Cette méthode estime les fréquences d'apparition des triplets de niveaux de gris  $(\ell_1, \ell_2, \ell_3)$ , au sein d'un cube de tri-occurrences. Le calcul consiste pour un couple d'orientations  $\theta = (\theta_1, \theta_2)$  et un couple de distances  $d = (d_1, d_2)$  à chercher les triplets de pixels  $\{x_i, x_j, x_k\}$  de niveaux de gris  $\{\ell_1, \ell_2, \ell_3\}$  de paramètres  $\theta$  et  $d$ .

Afin d'analyser le pouvoir discriminant des quinze attributs d'Haralick à l'ordre trois et d'en extraire les moins redondants, nous avons appliqué l'algorithme AP dans sa version non supervisée. L'indice de corrélation est utilisé comme mesure de similarité entre les attributs (cf. Tableau 6).

Le coefficient de corrélation entre deux attributs  $A_i$  et  $A_j$  de dimension  $N$ , s'écrit :

$$\text{corr}(A_i, A_j) = \frac{\text{cov}(A_i, A_j)}{\sqrt{\text{var}(A_i) \text{var}(A_j)}} \quad (5.1)$$

où  $\text{cov}(A_i, A_j)$  est la covariance entre les attributs  $A_i$  et  $A_j$ .

Les coefficients de corrélation entre les différents attributs sont à valeurs dans l'intervalle  $[-1, 1]$ . Le coefficient de corrélation  $\text{corr}(A_i, A_j)$  traduit la redondance entre les deux attributs  $A_i$  et  $A_j$ , celle-ci est d'autant plus forte que la valeur  $|\text{corr}(A_i, A_j)|$  est proche de 1.

Une étude approfondie sur la corrélation entre les attributs d'Haralick par la méthode de propagation non supervisée sur les images présentées dans 2.2.2 a permis de dégager les conclusions suivantes :

- les attributs du cube de tri-occurrences sont relativement corrélés,
- huit attributs sur 15 ont été sélectionnés par l'AP comme étant les plus représentatifs des données.

Nous calculons les huit attributs retenus à partir du cube de tri-occurrences noté  $M$  avec un nombre de niveaux de gris  $N_\ell$ .

Soit  $M_{d,\theta}(\ell_1, \ell_2, \ell_3)$  la fréquence d'apparition normalisée d'un triplet de niveaux de gris  $(\ell_1, \ell_2, \ell_3)$  distants successivement de  $d_1$  et  $d_2$  pour deux orientations  $\theta_1$  et  $\theta_2$ .

Les huit attributs du cube de tri-occurrences retenus sont :

**Contraste :**

$$Cont = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} [(\ell_1 - \ell_2)^2 (\ell_2 - \ell_3)^2 (\ell_1 - \ell_3)^2] M(\ell_1, \ell_2, \ell_3) \quad (5.2)$$

**Corrélation :**

$$Corr = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} \frac{(\ell_1 - \mu_i)(\ell_2 - \mu_j)(\ell_3 - \mu_k)}{\sigma_i \sigma_j \sigma_k} M(\ell_1, \ell_2, \ell_3) \quad (5.3)$$

où  $\mu_i, \mu_j, \mu_k, \sigma_i, \sigma_j$  et  $\sigma_k$  sont respectivement les centres de gravité et les écarts types de  $M_i, M_j$  et  $M_k$ , avec :

$$M_i = \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} M(\ell_1, \ell_2, \ell_3), \quad M_j = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} M(\ell_1, \ell_2, \ell_3) \text{ et } M_k = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M(\ell_1, \ell_2, \ell_3)$$

**Moment des différences inverses :**

$$IDM = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} \frac{M(\ell_1, \ell_2, \ell_3)}{1 + (\ell_1 - \ell_2)^2 + (\ell_1 - \ell_3)^2 + (\ell_2 - \ell_3)^2} \quad (5.5)$$

**Moyennes des sommes :**

$$SA = \sum_{h=3}^{3N_\ell} h M_i M_j M_k(h) \quad (5.6)$$

$$\text{avec : } M_i M_j M_k(h) = \sum_{\ell_1+\ell_2+\ell_3+3=h} M(\ell_1, \ell_2, \ell_3)$$

**Variances des sommes :**

$$SVAR = \sum_{h=3}^{3N_\ell} (h - SA)^2 M_i M_j M_k(h) \quad (5.7)$$

**Entropie :**

$$Entropie = - \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} M(\ell_1, \ell_2, \ell_3) \log(M(\ell_1, \ell_2, \ell_3)) \quad (5.8)$$

**Nuances du groupe :**

$$NG = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} (\ell_1 - F_i + \ell_2 - F_j + \ell_3 - F_k)^3 M(\ell_1, \ell_2, \ell_3) \quad (5.9)$$

**Proéminence du groupe :**

$$PG = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \sum_{\ell_3=0}^{N_\ell-1} (\ell_1 - F_i + \ell_2 - F_j + \ell_3 - F_k)^4 M(\ell_1, \ell_2, \ell_3) \quad (5.10)$$

Nous présentons ci-après les résultats d'évaluation de ces attributs pour le partitionnement des images monochromes. La méthode de classification utilisée ici est la version classique non supervisée de la propagation d'affinité.

### 5.2.2. Validation

Pour valider la méthode proposée de calcul des attributs de tri-occurrences, nous l'avons testée sur une base de données d'images synthétiques monochromes extraites de l'album de Brodatz [42]. Cette base sélectionnée est composée de 16 images dont chacune est constituée de cinq régions (trois uniformes et deux texturées). Les régions uniformes sont synthétisées suivant des modèles gaussiens avec des moyennes différentes et des écarts-types identiques.

Nous précisons que pour l'ensemble des expérimentations, le paramètre de préférence de l'AP est fixé à la valeur médiane de la matrice de similarité hors les éléments de la diagonale ; la valeur du paramètre de régularisation  $\alpha$  est fixée par défaut à 0,9 afin d'éviter les oscillations [5], et la métrique utilisée est L1.

Nous montrons dans la Figure 22) les résultats de classification en fonction de la nature des attributs extraits sur trois images. Les performances de trois types d'attributs sont comparées :

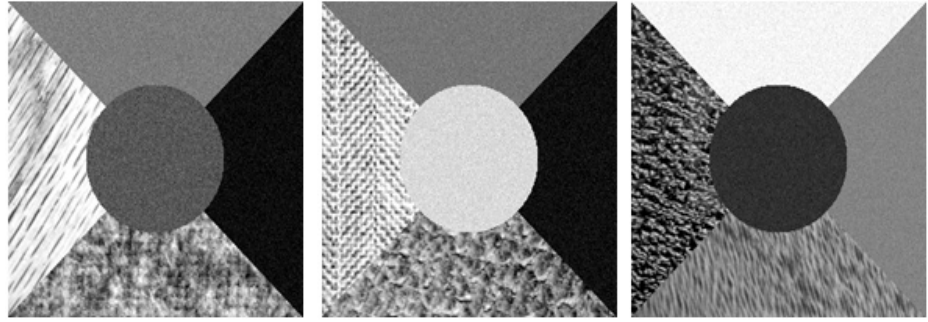
- Attributs de cooccurrences (8 attributs : cf. section 2.2.2),
- Attributs de tri-occurrences (8 attributs : cf. section 5.2.1),
- Attributs de tri-occurrences et moyenne (9 attributs).

Pour le calcul de ces attributs, une fenêtre de taille 3x3 a été utilisée. Pour chaque expérimentation, nous précisons le nombre de classes  $N_c$  estimé par l'AP ainsi que le taux de bonne classification moyenné sur l'ensemble des classes, noté ACCR (Average Correct



Classification Rate). La Figure 23 précise les performances obtenues pour chaque image en considérant cinq configurations liées aux types d'attributs : statistiques du premier ordre (quatre), ceux issus des ondelettes (deux), ceux issus de la cooccurrence (huit), ceux de la tri-occurrence (huit) et ceux de la tri-occurrence associés à la moyenne (neuf).

Pour l'ensemble des 16 images testées, les résultats expérimentaux confirment ceux obtenus sur chacune des images, à savoir que les attributs de tri-occurrences donnent de meilleurs résultats (ACCR : 75.04%) par rapport aux attributs statistiques du premier ordre (ACCR : 66.32%), ceux issus de la transformée en ondelettes (ACCR : 65.12%) et les attributs de la matrice de cooccurrences (ACCR : 68.40%). L'association de la moyenne avec les attributs de tri-occurrences a permis l'amélioration du taux de classification de ces derniers de 8.52% en moyenne. En effet, différentes combinaisons d'attributs ont été testées, seule la combinaison de la moyenne avec les attributs de tri-occurrences permet d'améliorer le taux de classification.

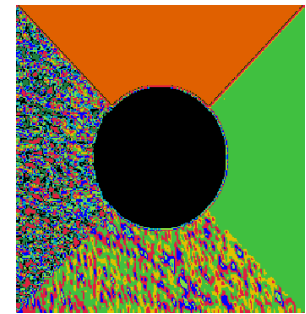
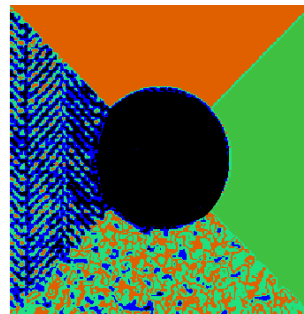
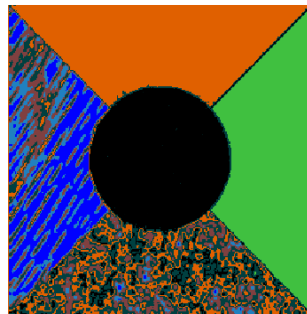


Exemples de trois images synthétiques à partitionner

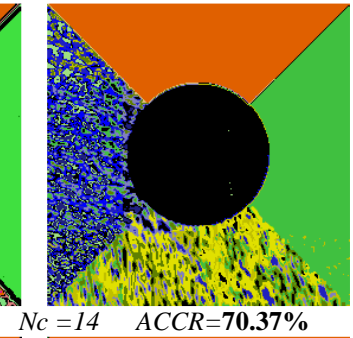
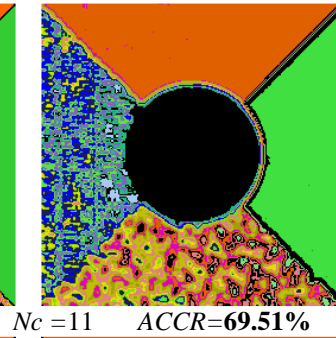
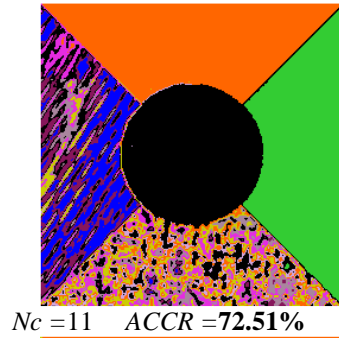
**Attributs utilisés**

**Résultats de classification**

Cooccurrences



Tri-occurrences



Tri-occurrences et  
moyenne

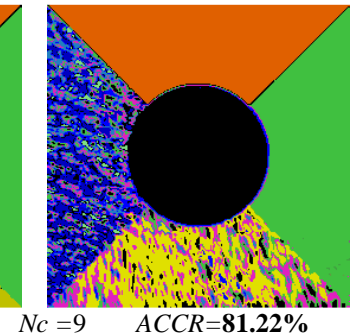
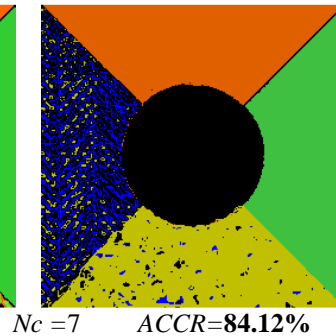
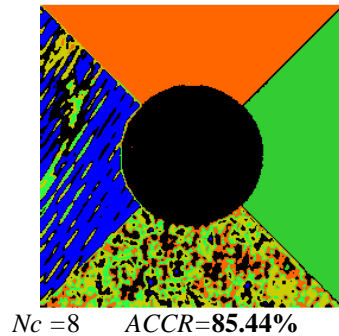


Figure 22: Résultats de partitionnement de trois images synthétiques avec différentes configurations d'attributs en utilisant l'AP standard.

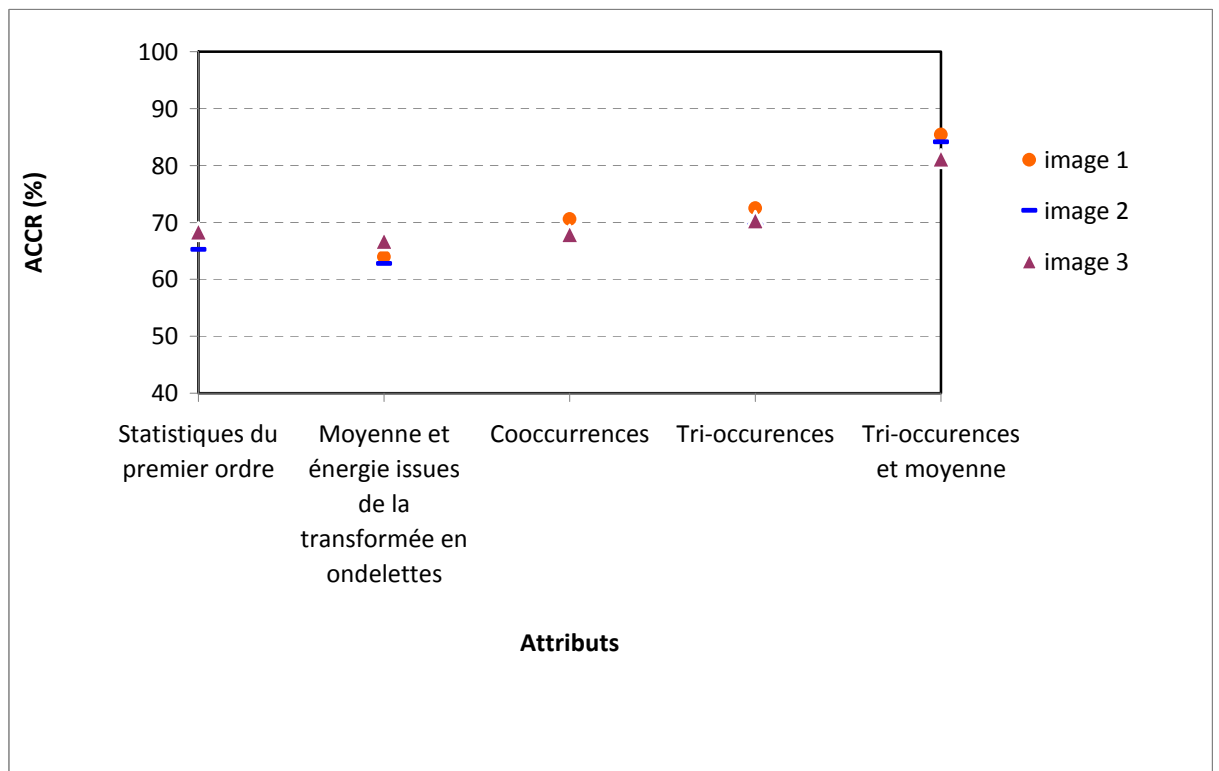


Figure 23: Performances des résultats de partitionnement des trois images synthétiques avec différentes configurations d'attributs.

Pour confirmer la supériorité de l'AP annoncée dans les différents travaux de la littérature, nous avons comparé ses performances à celles des méthodes de classification semi-supervisées, C-moyennes, FCM et ISODATA en fixant pour l'ensemble des méthodes le nombre de classes à 5 et en utilisant uniquement comme attributs ceux des tri-occurrences. Pour pouvoir comparer les résultats de l'AP à ceux des autres méthodes, nous l'avons utilisée dans sa version semi-supervisée. Le taux moyen de l'AP est de 95% contre 64.56% pour C-moyennes, 79.69% pour FCM et 63.52% pour ISODATA. Ces résultats confirment bien la supériorité de l'AP par rapport aux méthodes semi-supervisées souvent citées dans la littérature comme méthodes de référence. Il est à noter que l'AP donne de meilleurs résultats dans sa version semi-supervisée (nombre de classes égal 5) (95%) par rapport à la version non supervisée (75.04%), soit une amélioration de près de 20%. Cette différence est due au sur-partitionnement dans le cas de l'AP non supervisée.

### 5.2.3. Conclusion

Les différents résultats de classification présentés dans la section précédente nous ont donc permis d'étudier et d'évaluer les différents attributs pour dégager les conclusions suivantes :

- Le résultat de classification est fortement variable en fonction du choix des attributs,
- Les attributs de tri-occurrences donnent les meilleurs résultats de classification,
- En combinant la moyenne avec les attributs de tri-occurrences, le résultat de classification est amélioré pour l'ensemble des images testées,
- La méthode de propagation d'affinité non supervisée donne un sur-partitionnement de l'image. On peut remarquer cependant que le résultat de classification obtenu est visuellement cohérent bien que les textures supposées appartenir à la même classe ne soient pas parfaitement homogènes (connaissance *a priori* biaisée).

#### 5.2.4. Attributs de tri-occurrences spatiaux et spectraux

Pour les images multi et hyperspectrales, les attributs de tri-occurrences définis dans la section précédente peuvent être calculés suivant les dimensions spatiales en procédant bande par bande. Toutefois, cette méthode de calcul ne permet pas d'intégrer l'information spectrale, et traite les données comme s'il n'y avait aucune relation entre les bandes spectrales. La même remarque peut être faite si les attributs sont calculés uniquement suivant la dimension spectrale, sans utiliser les dimensions spatiales.

La Figure 24 montre la disposition des fenêtres d'analyses suivant trois dimensions.

- *Dimension spatiale* : les attributs sont calculés bande par bande sur des fenêtres définies dans le plan spatial.
- *Dimension spectrale* : les attributs sont calculés sur des bandes adjacentes.
- *Dimension spatiale et spectrale* : les attributs sont calculés sur des cubes, pour tenir compte à la fois de l'information spatiale et spectrale.

Pour mieux caractériser un pixel, nous avons donc intégré des attributs calculés suivant les dimensions spatiale et spectrale. Dans ce cas, le calcul des attributs de tri-occurrences pour les images multi ou hyperspectrales est réalisé sur un cube de taille impaire centré sur le pixel à caractériser en tenant compte de toutes les configurations des triplets de pixels (voxels) intégrant à chaque fois le pixel central comme l'illustre avec quelques configurations la Figure 25.

Avec ce procédé, chaque pixel pour une bande donnée peut être caractérisé par  $Q$  paramètres calculés de la même manière que les attributs présentés dans la section précédente. Au final chaque pixel du plan spatial sera représenté au total par un vecteur d'attributs dont la

dimension est  $Q \times N_b$  avec  $N_b$  le nombre de bandes spectrales. Dans le cas où seulement trois bandes sont considérées suivant la dimension spectrale le vecteur d'attributs pour caractériser un pixel de la bande centrale a une dimension de  $81 \times N_b$

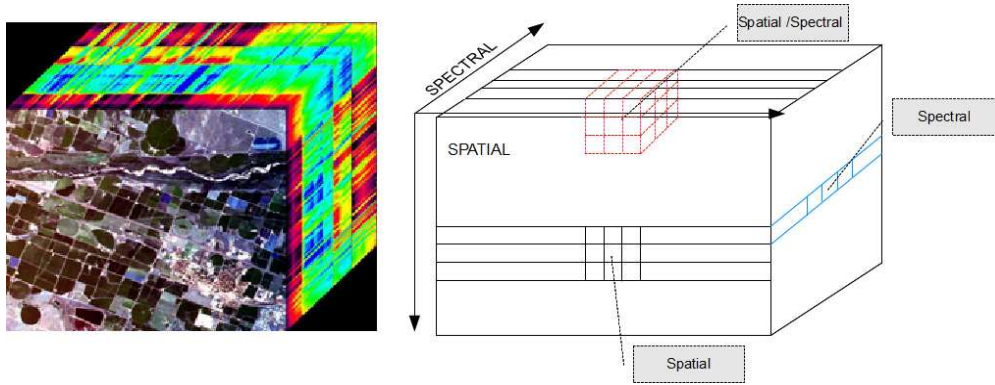


Figure 24: Différentes dispositions des fenêtres d'analyses pour l'extraction d'attributs

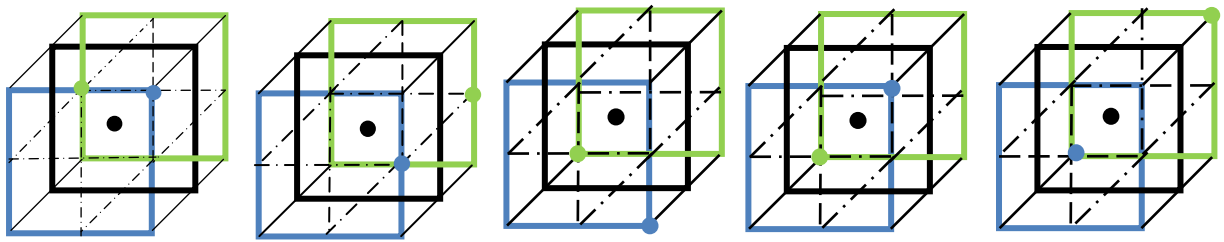


Figure 25: Quelques exemples de configurations de tri-occurrences dans un cube 3x3x3

Nous allons évaluer dans la section suivante la pertinence des attributs spatiaux et spectraux dans le cas des images hyperspectrales.

### 5.2.5. Validation

Une première évaluation de l'étape d'extraction d'attributs de tri-occurrences spatiaux et spectraux a été réalisée sur l'image hyperspectrale déjà présentée dans la section 4.3 du chapitre 4. Dans cette première expérimentation, nous rappelons que la taille choisie pour l'image construite est limitée à  $64 \times 64$  pixels pour deux raisons : la première est la possibilité d'appliquer l'AP sur un nombre d'individus compatible avec les capacités de traitement classique en environnement Matlab, la seconde est la limitation des données de la vérité de terrain disponible sur la zone couverte par l'image acquise permettant la validation.

La Figure 26 montre les résultats de classification de l'AP en mode non supervisé de l'image montrée plus haut en Figure 15, avec les mêmes configurations d'attributs que dans la section 2.2.2. Le résultat de la classification en considérant uniquement la signature spectrale des pixels est également donné.

Les remarques évoquées dans la section 2.2.2 restent valables ici. En effet, les tri-occurrences offrent les meilleures performances, ainsi que leur combinaison avec la moyenne, qui améliore encore la qualité de classification de 11.2 %. La combinaison du moment du premier ordre (moyenne) avec les attributs d'ordre supérieur permet de caractériser des structures de texture de différents types : fines, grossières, peu texturées ou fortement texturées. Ce qui justifie l'amélioration du taux de bonne classification.

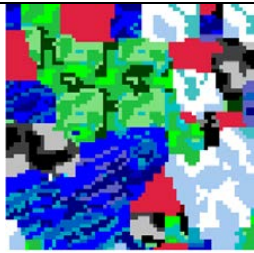
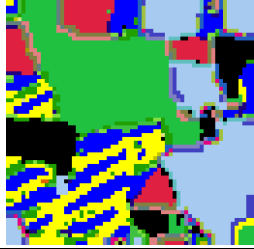

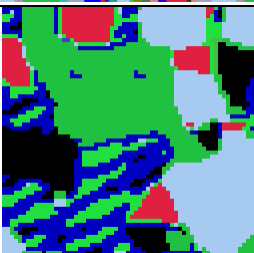

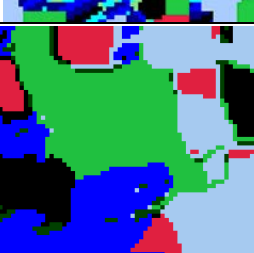

Attributs utilisés	Résultats de classification	
Signatures spectrales		$N_c = 19$ $ACCR = 77.44\%$
Statistiques du premier ordre		$N_c = 20$ $ACCR = 71.65\%$
Moyenne et énergie issues de la transformée en ondelettes		$N_c = 35$ $ACCR = 65.47\%$
Cooccurrence		$N_c = 6$ $ACCR = 80.16\%$
Tri-occurrences		$N_c = 8$ $ACCR = 82.15\%$
Tri-occurrences et moyenne		$N_c = 6$ $ACCR = 93.35\%$
		

Figure 26: Résultats de partitionnement de l'image hyperspectrale synthétique avec différentes configurations d'attributs

### 5.2.6. Conclusion

Les résultats présentés ici ont montré que :

- Les attributs de tri-occurrences proposés permettent une meilleure caractérisation vis à vis des attributs statistiques d'ordre inférieur,
- La surestimation du nombre de classes par la propagation d'affinité est encore confirmée pour l'image hyperspectrale testée.
- L'amélioration du résultat de classification par la combinaison de la moyenne et les huit attributs de tri-occurrences reste vraie pour l'image hyperspectrale.

Les attributs de tri-occurrences associés à la moyenne sont donc retenus pour le partitionnement des images hyperspectrales dans la suite de ce travail.

Il est à noter que nous nous sommes limités dans cette section aux images hyperspectrales de petite taille spatiale afin de pouvoir appliquer la méthode de propagation d'affinité.

## 5.3. Réduction de dimension et classification non supervisée par propagation d'affinité

Afin de pouvoir appliquer la méthode de propagation d'affinité pour le partitionnement des données de grandes dimensions spatiales comme les images hyperspectrales (grand nombre de pixels), nous proposons ici une étape qui consiste à réduire le nombre d'individus avant l'application de l'AP. La section suivante précise la démarche suivie.

### 5.3.1. Réduction du nombre d'individus à classer

Nous introduisons ici une étape préliminaire à la classification par la méthode AP, dont l'objectif est la réduction de la taille de la matrice de similarité ce qui doit permettre d'appliquer cette méthode sur des images de grandes dimensions spatiales. Cette étape consiste à regrouper automatiquement les individus ayant une très forte similarité pour ne pas affecter la qualité des données à classer par l'AP, et à remplacer chaque groupe homogène formé par un seul représentant. Le critère d'agrégation utilisé est celui de la distance  $L_1$ . Pour réduire le temps de calcul, l'image est divisée en blocs carrés de taille  $N_{bc}$  individus, puis la recherche de similarité est réalisée de manière parallèle sur chacun des blocs.



La procédure de réduction est réalisée de manière itérative en exploitant seulement la signature spectrale des pixels. Plus précisément, les pixels dont les spectres sont identiques sont d'abord regroupés lors de la première itération. À ce niveau, chaque sous-groupe formé est représenté par un individu choisi aléatoirement parmi eux puisque les signatures spectrales sont identiques. Ensuite, à partir de la deuxième itération, le pixel ayant la plus petite distance au centre de gravité de la sous-classe est sélectionné pour la représenter. Puis, à chaque itération, l'appariement est réalisé sur l'ensemble des pixels retenus à l'itération précédente en relâchant la contrainte sur le critère de similarité. La valeur du critère de similarité pour l'arrêt de l'appariement est estimée localement pour chaque bloc. Pour  $N$  individus à classer, cette étape regroupe chaque individu  $x_i$  avec l'individu  $x_k$  présentant la distance minimale par rapport à tout le reste de l'ensemble. Si en parcourant le reste des données, il existe un individu  $x_j$  ayant une distance plus faible par rapport à un individu  $x_k$  que celle d'un individu  $x_i$ , le lien entre les individus  $x_i$  et  $x_k$  est rompu ( $x_i$  sera seul) et  $x_k$  sera regroupé avec  $x_j$ . La procédure par bloc ne nécessite pas la construction d'une matrice de similarité, mais seulement des matrices de distance qui ont les mêmes tailles que celles des blocs de l'image à partitionner. La procédure de réduction est automatisée car à la fin de chaque passage de l'algorithme, la valeur de la distance maximale d'agrégation à ne pas dépasser est estimée. Cette valeur notée  $\delta_B$  est estimée pour chacun des blocs comme suit :

Soient  $N_{bc}$  le nombre d'individus d'un bloc  $B$  de l'image et  $Sim_B$  un ensemble de taille  $((N_{bc}^2 - N_{bc})/2)$  représentant les similarités  $S_m$  entre chaque couple d'individus dans ce bloc.

L'indice de similarité  $S_m$  entre le couple d'individus  $x_i$  et  $x_k$  s'écrit :

$$S_m(x_i, x_k) = \|a_i - a_k\|_1 = \sum_{u=1}^{N_a} |a_{i_u} - a_{k_u}| \quad (5.11)$$

avec  $N_a$  le nombre d'attributs et  $a_i$  le vecteur d'attributs représentant l'individu  $x_i$ .

La valeur limite d'agrégation est estimée comme suit :

- Calcul de la valeur moyenne  $\mu_B$  et de la valeur de l'écart type  $\sigma_B$  de l'ensemble  $Sim_B$ .
- Calcul du nouveau sous-ensemble  $Sim_B' \subset Sim_B$  ayant des valeurs dans l'intervalle  $[\mu_B - \sigma_B, \mu_B + \sigma_B]$ .
- Estimation de la valeur de l'écart-type du nouveau ensemble  $Sim_B'$  donnant le seuil d'agrégation  $\delta_B$  du bloc  $B$ .

### 5.3.2. Validation de l'étape de réduction

Pour montrer la stabilité de la méthode de réduction proposée par rapport à la variation de la taille des blocs, nous avons exécuté l'approche développée pour différentes tailles en calculant à chaque fois le taux de classification correspondant donné par l'AP sur l'ensemble des pixels retenus dans les différents blocs. Ces résultats sont également comparés au taux de classification de référence obtenu sur l'image sans découpage.

Les différentes étapes de la procédure de validation sont les suivantes :

1. Diviser l'image en blocs,
2. Appliquer la réduction sur chacun des blocs,
3. Calculer les attributs de tri-occurrence et la moyenne,
4. Appliquer la méthode de classification par AP semi-supervisée sur les pixels retenus des différents blocs en fixant le nombre de classes à 5,
5. Calculer le taux de classification.

Pour valider l'étape de réduction (2) nous avons utilisé l'image déjà présentée dans la section 4.3. (Figure 15) Quatre tailles de blocs sont utilisées comme le montre la Figure 27 ; un bloc de taille 64x64 (image entière), 4 blocs de taille 32x32 pixels, 16 blocs de tailles 16x16 pixels et 64 blocs de 8x8 pixels. Dans cette figure, les résultats de l'approche proposée de réduction automatique par bloc sont donnés en indiquant le nombre de pixels retenus ( $N_r$ ). À titre de comparaison, les résultats de réduction par l'AP de manière non supervisée sont donnés. L'AP a été utilisée dans l'étape de réduction car elle surestime le nombre de classes et par conséquent, permet de former une première partition avec des classes fortement homogènes. Ces résultats montrent que le nombre de pixels retenus par l'approche proposée est toujours très supérieur à celui donné par l'AP en mode non supervisé. Il est à rappeler que seules les signatures spectrales sont utilisées dans cette expérimentation comme attributs dans l'étape de réduction des pixels.

Les taux de bonne classification obtenus (ACCR %) après l'application de l'AP en mode supervisé sur l'ensemble des pixels retenus par les deux approches de réduction en fonction de la taille des blocs de découpage de l'image originale sont donnés dans la Figure 28. Le Tableau 8 donne les détails de ces performances après l'introduction de l'étape de réduction sur des blocs 64x64 (l'image entière), 32x32, 16x16 et 8x8. Les taux obtenus après réduction

avec l'approche développée sont élevés ; de plus, ils sont pratiquement indépendants de la taille des blocs. Par contre, l'application de l'AP en mode non supervisé à la fois dans l'étape de réduction, et dans celle de la classification donne des résultats moins probants.

Ces résultats montrent donc que le taux de bonne classification peut diminuer si la phase de réduction n'est pas optimisée.

L'application de l'étape de réduction proposée, associée à l'AP semi-supervisée sur toute l'image donne un taux moyen de bonne classification de 97.51%. Ce taux est supérieur à celui obtenu par l'application directe de l'AP semi-supervisée sans réduction sur toute l'image en fixant le nombre de classes à 5 (ACCR : 90.72%). En effet, le fait de remplacer, dans l'étape de réduction, chaque groupe d'individus strictement similaires par un seul représentant avant l'étape de classification, perturbe moins le calcul des poids affectés aux individus par l'AP et par conséquent fausse moins les résultats de partitionnement.

Dans cette évaluation, nous avons examiné également la validité de l'approche de réduction développée par appariement automatique suivant la métrique L1, en introduisant des attributs différents dans l'étape de classification par AP. La Figure 29 montre ces résultats. Nous observons que quel que soit le type d'attributs, la phase de réduction améliore les performances de l'AP, ce qui rend cette étape préliminaire obligatoire avant l'application de l'AP. Il est à noter que les résultats obtenus avec les attributs de la tri-occurrence associés à la moyenne sont les plus stables.

En conclusion, les différents résultats de l'approche de réduction proposée sont pertinents ; de plus ils sont stables par rapport à la taille des blocs sur lesquels elle est appliquée. Pour appuyer cette conclusion, les résultats de cette étape de réduction ont été également comparés à ceux de l'algorithme neuronal SOM (Self-Organizing Map) [82] souvent utilisé comme étape préliminaire avant celle de la classification comme nous l'avons mentionné dans la section 1.2.2.3 . La Figure 30 montre les résultats de l'application de l'AP en fixant le nombre de classes à 5, sans étape de réduction préliminaire (a), avec une réduction par notre approche (b) et avec réduction par le SOM (c). Notre approche donne le meilleur taux moyen de bonne classification (97.51 % contre 95.36%). Le SOM est utilisé ici en mode semi-supervisé, où le nombre de pixels est le même que celui donné par notre étape de réduction. Le résultat se dégrade encore si le SOM est utilisé en mode non supervisé

(74,32%). Nous avons également utilisé dans cette expérimentation les signatures spectrales des pixels comme attributs dans l'étape de réduction.


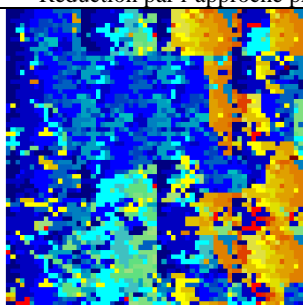
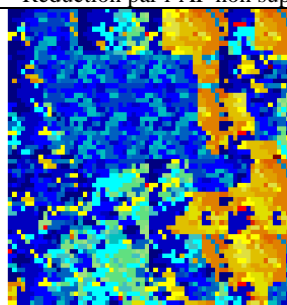
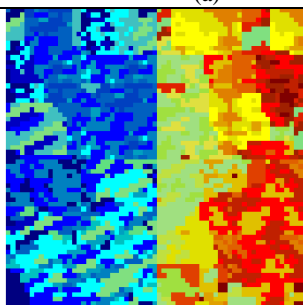
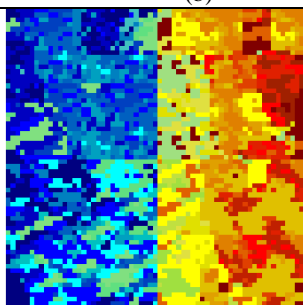
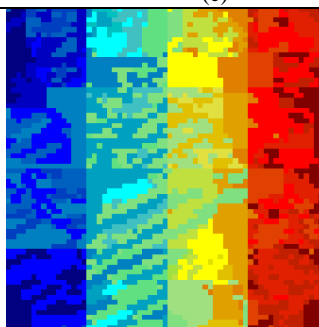
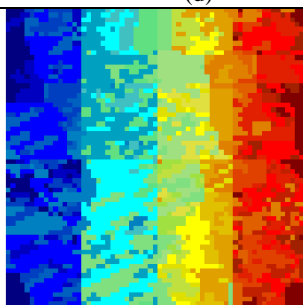
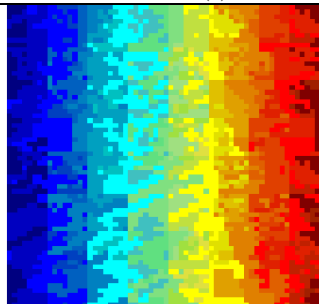
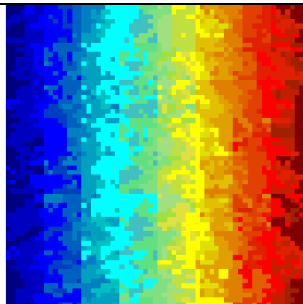
 Image originale		
	Réduction par l'approche proposée	Réduction par l'AP non supervisée
Bloc 64x64	 (a) $N_r=2281$	 (b) $N_r=214$
Blocs 32x32 pixels	 (c) $N_r=2154$	 (d) $N_r=110$
Blocs 16x16 pixels	 (e) $N_r=2023$	 (f) $N_r=225$
Blocs 8x8 pixels	 (g) $N_r=1962$	 (h) $N_r=448$

Figure 27: Réduction du nombre de pixels à classer par appariement en fonction de la taille des blocs par deux approches.

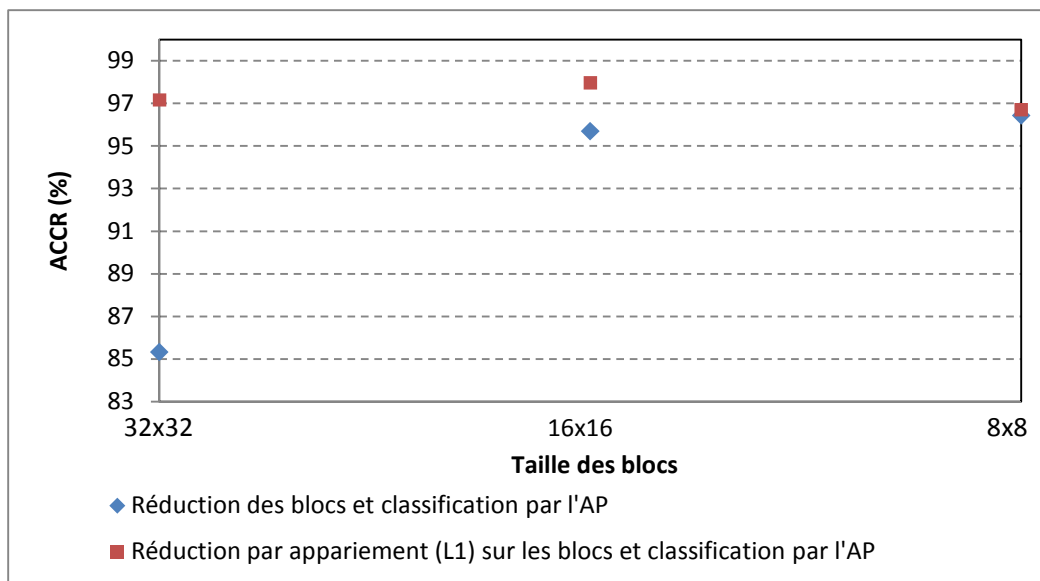


Figure 28: Performances de l'approche de réduction et de classification en fonction de la taille des blocs de découpage de l'image originale.

Tableau 8: Résultats de classification en fonction de la taille des blocs

Taille de blocs (population totale 4096 pixels)	Nombre d'individus retenus après réduction par l'approche développée	Taux de classification par l'AP en mode semi- supervisé	Nombre d'individus retenus après réduction par l'AP	Taux de classification par l'AP en mode semi- supervisé
8x8	1962	<b>96.52%</b>	448	96.43%
16x16	2023	<b>97.47%</b>	225	95.70%
32x32	2154	<b>97.49%</b>	110	85.32%
<b>64x64</b>	<b>2281</b>	<b>97.51%</b>	214	90.72%
64x64	sans réduction	90.72%	sans réduction	90.72%

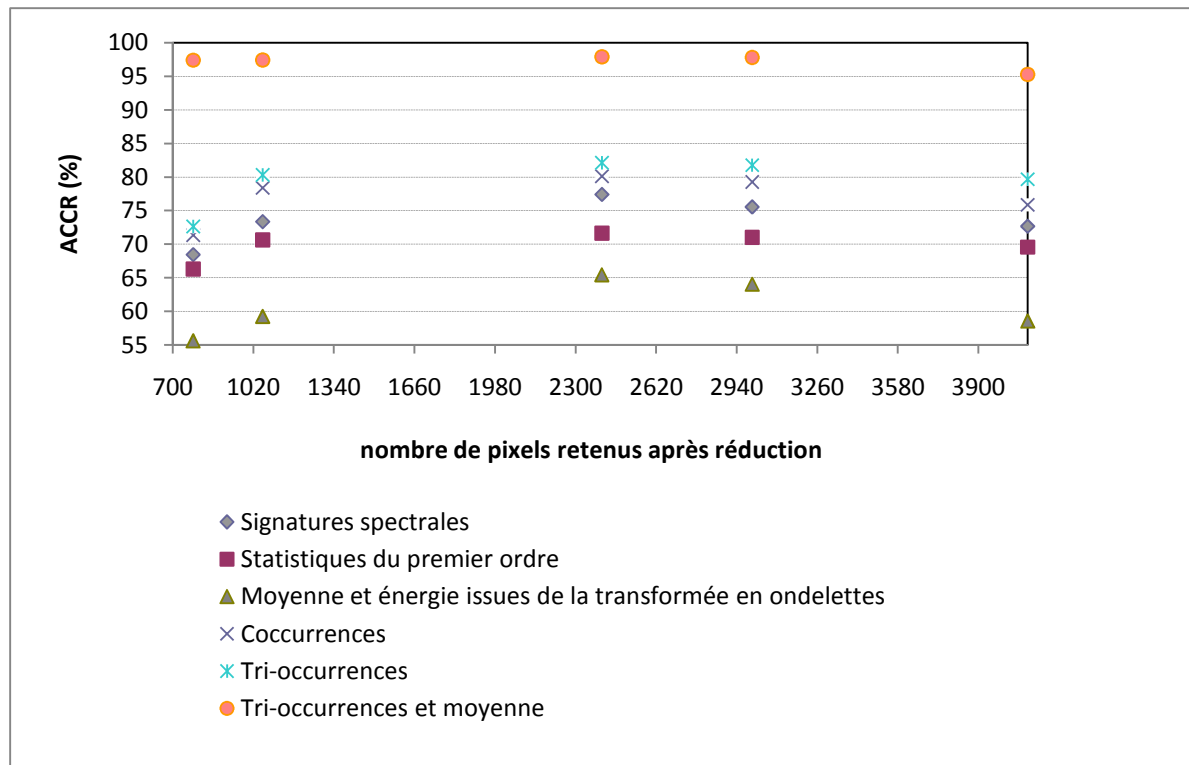


Figure 29: Performance de l'AP (ACCR (%)) en fonction du nombre de pixels après réduction par appariement (L1) et par type d'attributs.

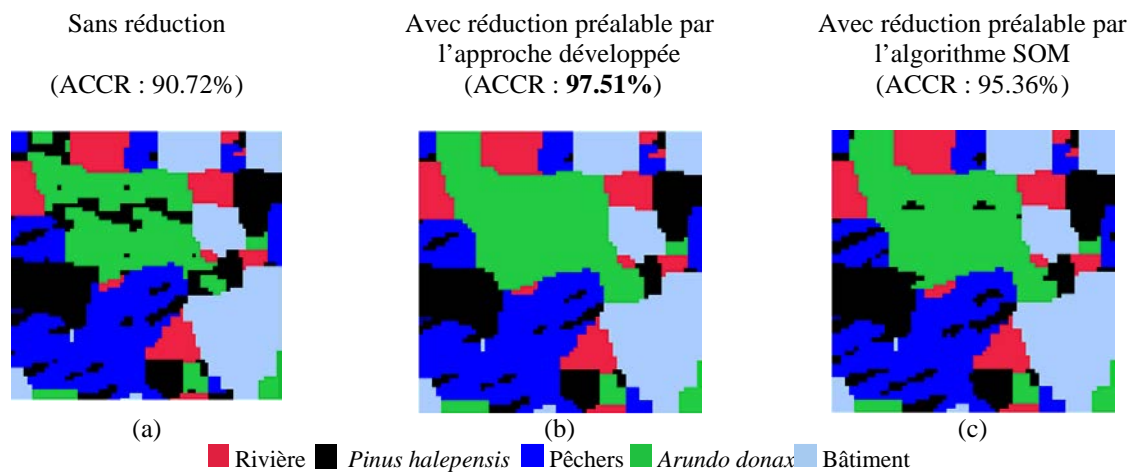


Figure 30 : Résultats de classification de l'AP (nombre de classes fixé à 5) avec et sans réduction.

Pour montrer à nouveau l'intérêt de l'introduction de l'étape de réduction du nombre d'individus avant l'application de l'AP, la Figure 31 montre les résultats de classification obtenus par l'AP en mode non supervisé avec et sans l'étape préliminaire de réduction sur l'image hyperspectrale construite. En fixant le paramètre de préférence  $p$  à la valeur médiane de la matrice de similarité (mode non supervisé), le nombre estimé de classes par l'AP est de

19 sans réduction, alors qu'il est réduit à 9 après l'application de l'étape de réduction proposée sur 4 blocs. Après cette étape, le nombre de classes a été réduit, mais il n'est toujours pas optimal car l'image est sur-segmentée.

Nous proposons dans la section suivante une approche pour optimiser l'estimation du nombre de classes.

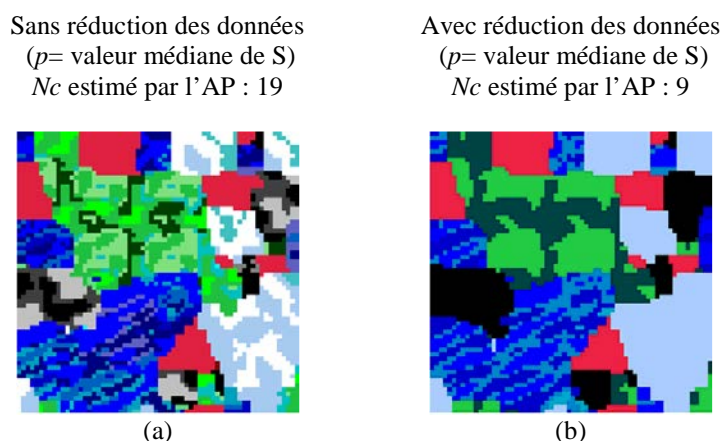


Figure 31 : Résultat de classification de l'AP standard en mode non supervisé sans et avec réduction.

### 5.3.3 Classification avec optimisation de l'estimation du nombre de classes

Comme nous l'avons indiqué dans les sections précédentes, le nombre de classes déterminé par l'AP est surestimé lorsque le paramètre de préférence est choisi comme la médiane de la matrice de similarité  $S$ . En effet, ce paramètre  $p$  qui traduit la force d'attractivité propre à chaque individu, contrôle implicitement le nombre d'exemplaires représentant chaque classe. Plus sa valeur est faible, plus l'individu cède sa priorité à devenir représentant de classe. Par contre, plus sa valeur est grande (proche de 0) plus il va être privilégié pour être représentant de sa classe. En adoptant comme valeur de  $p$  celle de la médiane des valeurs de  $S$  (sans tenir compte des éléments de la diagonale) comme préconisé dans [5], on observe que l'AP génère un nombre de classes important, et par conséquent une mauvaise estimation du nombre de classes, ce qui se traduit par un taux d'erreur de classification élevé. Par conséquent, l'estimation du nombre de classes, passe nécessairement par une bonne estimation du paramètre de préférence  $p$ .

Pour estimer la valeur du paramètre  $p$  donnant la meilleure partition et donc une estimation optimisée du nombre de classes, nous avons introduit le critère d'évaluation ( $EC$ ) lié à la variance interclasse défini par Levine et Nazif [140] et discuté dans la première partie



de ce mémoire (Chapitre 4). Rappelons que la meilleure partition retenue est celle qui maximise le critère  $EC$ .

Les différentes étapes pour l'estimation du paramètre  $p$  donnant la meilleure partition sont :

- 1 : Exécution de l'AP avec  $p$ = valeur médiane de  $S$ ,
- 2 : Calcul des valeurs de la matrice de responsabilité  $R$  de la dernière itération notée  $R_f$ ,
- 3 : Recherche dichotomique de la valeur de  $p$  qui maximise  $EC$  dans l'intervalle  $[p_{inf}, p_{sup}]$ , avec  $p_{inf} = \min(R_f)$  et  $p_{sup} = \text{médiane}(R_f)$ .

En appliquant cette approche d'estimation du nombre de classes, la meilleure partition après l'étape de réduction, en n'utilisant que les signatures spectrales comme attributs, puis l'AP avec les attributs de tri-occurrence et la moyenne, est obtenue pour 5 classes. La Figure 32 montre l'évolution des résultats intermédiaires d'estimation du nombre de classes (recherche de la valeur de  $p$ ) en fonction de l'évolution du critère de validation. Le taux de bonne classification correspondant au nombre de classes correctement estimé (figure 32 (c)) est de 97.51%. Le Tableau 9 montre la matrice de confusion correspondant à ce résultat (5 classes).

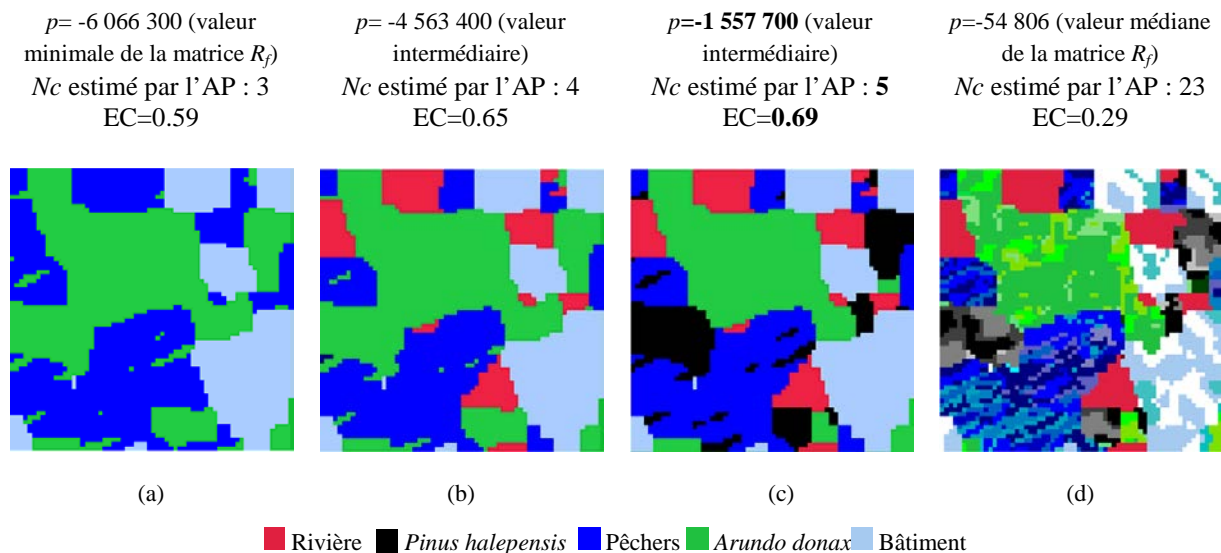


Figure 32 : Partitionnement de l'image hyperspectrale synthétique par l'AP avec étape de réduction du nombre d'individus : évolution du nombre de classes en fonction de la valeur du paramètre  $p$ .

Tableau 9 : Matrice de confusion (5 classes) correspondant au résultat de partitionnement de la Figure 32(c), CCR en %, (.) : nombre de pixels.

Classes prédites par l'approche proposée (nombre de pixels)	Classes de la vérité de terrain (nombre de pixels)				
	Rivière (452)	<i>Pinus halepensis</i> (500)	Pêcher (1189)	<i>Arundo donax</i> (1068)	Bâtiment (887)
Rivière (468)	<b>100%</b> (452)	0	1.35% (16)	0	0
<i>Pinus halepensis</i> (630)	0	<b>100%</b> (500)	10.93% (130)	0	0
Pêcher (1041)	0	0	<b>87.55%</b> (1041)	0	0
<i>Arundo donax</i> (1070)	0	0	0.17% (2)	<b>100%</b> (1068)	0
Bâtiment (887)	0	0	0	0	<b>100%</b> (887)

## 5.4. Conclusion

L'approche proposée de réduction et de classification avec estimation du nombre de classes a été validée sur l'image hyperspectrale. Les résultats ont montré la robustesse et la stabilité de la méthode en exploitant directement les signatures spectrales dans l'étape de réduction et les attributs de tri-occurrence associés à la moyenne dans la phase de classification par l'AP.

L'application de l'AP après une étape préliminaire de réduction du nombre d'individus sur des images hyperspectrales, spatialement de grande taille a été validée. De plus elle contribue à l'augmentation du taux de bonne classification de l'AP. Les résultats de la recherche automatique de la valeur optimale du paramètre  $p$  pour estimer le nombre de classes sont probants.

Pour valider la chaîne développée dans cette thèse, nous présentons dans le chapitre suivant les résultats de classification obtenus sur trois applications réelles qui entrent dans le cadre de la problématique de la gestion et du suivi de l'environnement à partir d'images aériennes hyperspectrales ou d'images multispectrale satellitaires. La première application porte sur la de détection des plantes invasives, et les deux dernières sur la classification du couvert paysager.

## Chapitre 6

# Application à l'identification de plantes invasives et du couvert paysager

### 6.1 Identification de plantes invasives

Les plantes invasives pour un territoire donné sont des plantes envahissantes étrangères à l'ensemble des plantes indigènes. Dans certains cas, la détection de ces plantes à une étape précoce est d'un grand intérêt sur le plan environnemental et économique. L'objectif d'une détection à un stade précoce permet d'entreprendre des actions de gestion ultérieures appropriées de façon à limiter leur développement.

Cette première application concerne donc l'identification de plusieurs plantes invasives et non-invasives d'une zone de la région de Murcia (sud est de l'Espagne). La disponibilité de relevés sur cette zone nous a permis de valider l'approche non supervisée développée dans le cadre de cette thèse.

Nous présentons dans ce chapitre les différents résultats de classification donnés par notre approche ainsi qu'une étude comparative avec d'autres méthodes de classification.

#### 6.1.1 Présentation des données

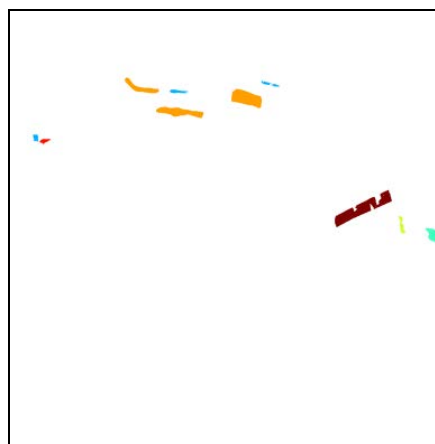
Nous avons choisi une zone de l'image hyperspectrale Cieza -Région de Murcia, Espagne- (8075x9748 pixels) qui contient le maximum de données de vérité de terrain. La taille de la zone choisie est de 1000x1000 pixels. Six classes sont présentes dans cette zone dont trois correspondent à des plantes invasives (*Phragmites australis*, *Tamarix* et *Arundo donax*), comme illustrées sur la Figure 33.

Zone de l'image hyperspectrale choisie visualisée en  
RVB  
(650, 550, 450) nm (taille 1000x1000)

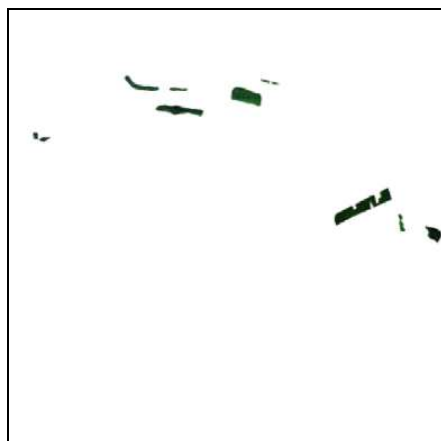
Masques des classes de la vérité de terrain



(a)



(b)



(c)

Zones de l'image hyperspectrale correspondant aux masques. Image visualisée en RVB

(650, 550, 450) nm

- |                                                                 |                                                              |                                                           |
|-----------------------------------------------------------------|--------------------------------------------------------------|-----------------------------------------------------------|
| <span style="color: red;">■</span> Pêcher                       | <span style="color: red;">■</span> <i>Tamarix</i>            | <span style="color: orange;">■</span> <i>Arundo donax</i> |
| <span style="color: blue;">■</span> <i>Phragmites australis</i> | <span style="color: green;">■</span> <i>Pinus halepensis</i> | <span style="color: yellow;">■</span> <i>Ulmus minor</i>  |

Figure 33 : Image hyperspectrale Cieza acquise par le capteur AISA Eagle et zones d'intérêt à partitionner.

### 6.1.2 Résultats d'identification

Cette section présente les résultats obtenus par l'approche développée pour la détection de trois types de plantes invasives (*Phragmites australis*, *Tamarix* et *Arundo donax*). Nous présentons dans un premier lieu les résultats de classification en exploitant directement les signatures spectrales de chaque pixel à la fois dans les étapes de réduction des

pixels et de classification. Dans un second lieu, nous présentons les résultats obtenus en utilisant les signatures spectrales dans l'étape de réduction, et les autres catégories d'attributs déjà évoqués dans le chapitre précédent en entrée du classifieur AP.

- ***Classification à partir des signatures spectrales***

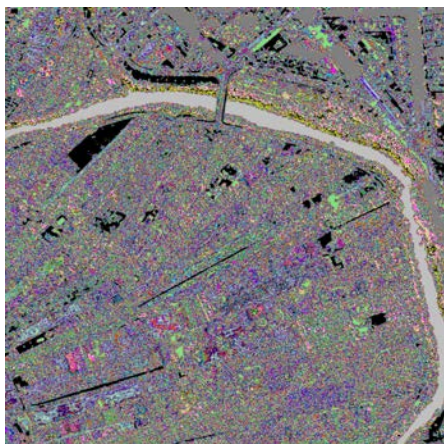
Les étapes de validation de l'approche proposée sont :

1. Division de l'image en blocs (400 blocs de taille 50 x 50 pixels),
2. Application de l'étape de réduction sur chacun des blocs,
3. Application de l'AP avec l'optimisation de l'estimation du nombre de classes,
4. Calcul du taux de bonne classification (ACCR).

L'image hyperspectrale de la Figure 34 montre les résultats intermédiaires de l'estimation du nombre de classes ( $N_c$ ), selon le critère d'évaluation  $E_C$  (recherche de la valeur optimale de  $p$ ) et les résultats de partitionnement correspondants. La première colonne donne le résultat de partitionnement sur l'ensemble de l'image, tandis que, pour une meilleure lisibilité, la deuxième colonne montre le résultat correspondant seulement aux zones d'intérêts.

Le meilleur résultat maximisant le critère  $E_C$  est donné pour huit classes. Le Tableau 10 donne la matrice de confusion correspondante où le taux de bonne classification par classe (Correct Classification Rate (CCR)) est indiqué. Les résultats de l'application de l'approche proposée donnent un taux de bonne classification de 97,66% pour les six classes de la vérité de terrain. Ces résultats montrent que deux des trois plantes invasives (*Tamarix* et *Arundo donax*) sont détectées à 100%. Le taux de bonne classification pour la classe *Phragmites australis* est de 86,69%.

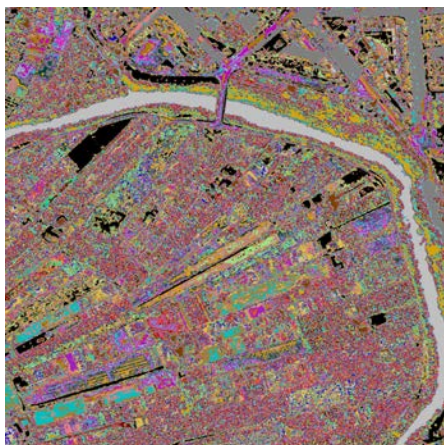
$p = -52\ 200$   
(valeur minimum de la matrice  $R_f$ )  
 $N_c$  estimé par l'AP : 313     $E_C = 0.31$



$N_c$  estimé par l'AP : 68     $E_C = 0.25$



$p = -28\ 500$  (valeur intermédiaire)  
 $N_c$  estimé par l'AP : 41     $E_C = 0.48$



$N_c$  estimé par l'AP : 26     $E_C = 0.50$



$p = -66$  (valeur médiane de la matrice  $R_f$ )  
 $N_c$  estimé par l'AP : 35     $E_C = 0.71$



$N_c$  estimé par l'AP : 8     $E_C = 0.68$



- *Arundo donax*
- *Tamarix*
- *Pinus halepensis*
- *Ulmus minor*
- *Phragmites australis*
- Pêcher

Figure 34 : Partitionnement de l'image hyperspectrale Cieza en fonction de la valeur du paramètre  $p$  : 1ère colonne: partitionnements de l'image complète ; 2ème colonne : Résultats de détection de plantes invasives et non invasives sur les zones d'intérêt.

Tableau 10: Matrice de confusion correspondant aux résultats de partitionnement à 8 classes.  
CCR en %, (.) : nombre de pixels.

Classes discriminées automatiquement par l'approche proposée (nombre de pixels)	Classes de la vérité de terrain (nombre de pixels)					
	<i>Phragmites australis</i> (556)	<i>Arundo donax</i> (4305)	<i>Tamarix</i> (162)	Pêcher (3115)	<i>Ulmus minor</i> (795)	<i>Pinus halepensis</i> (274)
<i>Phragmites australis</i> (482)	<b>86.69%</b> (482)	0	0	0	0	0
<i>Arundo donax</i> (4305)	0	<b>100%</b> (4305)	0	0	0	0
<i>Tamarix</i> (162)	0	0	<b>100%</b> (162)	0	0	0
Pêcher (3115)	0	0	0	<b>100%</b> (3115)	0	0
<i>Ulmus minor</i> (795)	0	0	0	0	<b>100%</b> (795)	0
<i>Pinus halepensis</i> (272)	0	0	0	0	0	<b>99.27%</b> (272)
Classe anonyme 1 (8)	1.43% (8)	0	0	0	0	0
Classe anonyme 2 (68)	11.87% (66)	0	0	0	0	0.72% (2)

Nous avons également comparé les performances de l'étape de réduction proposée à celle de l'approche de SOM semi-supervisée et non supervisée, sur la même image hyperspectrale (1000 x 1000 pixels x 62 bandes). Dans le cas de l'utilisation de l'algorithme SOM, en mode semi-supervisé (nécessité de la connaissance *a priori* du nombre de classes), nous avons estimé le nombre de classes par l'AP. Le Tableau 11 montre que les résultats de la réduction par SOM ne sont pas satisfaisants ni en mode semi-supervisé ni en mode non supervisé. En effet, les résultats de la classification dans les deux cas donnent respectivement un taux de classification de 47,20% et 45,60% contre 97,66% par notre approche. Ces résultats confirment les expérimentations obtenues sur l'image synthétique et montrent que la réduction des données par le SOM n'est pas satisfaisante.

Enfin, nous avons comparé les résultats de notre approche de classification avec ceux des méthodes classiques semi-supervisée C-moyennes et ISODATA souvent utilisées en télédétection. Pour être en mesure d'utiliser ces méthodes semi-supervisées, une connaissance *a priori* d'un certain nombre de paramètres est nécessaire : le nombre de classes, les valeurs de certains seuils et le nombre d'itérations. Nous précisons que ces deux méthodes ont été appliquées sur les données après l'étape de réduction réalisée par notre méthode. Nous

estimons d'abord le nombre de classes  $N_c$  par notre approche, puis nous introduisons cette valeur comme connaissance *a priori* pour les deux autres. Pour l'algorithme C-moyennes, le meilleur taux de classification (65.03%) a été obtenu avec des valeurs de seuil fixées à 5% et trois itérations et, pour ISODATA, le meilleur taux (62.81%) a été obtenu avec un seuil de 5%, trois itérations, un nombre minimum de pixels par classe égale à 1, une variance maximum de 1, une distance minimale entre les classes de 5 et un nombre maximum de paires de classes pouvant être fusionnées égale à 2.

Le taux de bonne classification obtenu par notre approche est de **97.66%**.

Les résultats des expérimentations confirment l'efficacité de notre approche par rapport aux autres méthodes comparées à la fois pour l'étape de réduction (par rapport à SOM) et pour l'étape de classification (par rapport aux C-moyennes et ISODATA).

Tableau 11: Comparaison des performances de trois classifieurs sur la zone d'image Cieza (1000x1000 pixels, 62 bandes spectrales).

	Approche proposée (étape de réduction + AP modifié)	SOM (semi-supervisée) + AP modifié	SOM (non supervisée) +AP modifié
Nombre initial de pixels à classer	1 000 000		
Nombre de pixels après l'étape de réduction	45 518	45 518	4943
$N_c$ estimé sur toute l'image	35	35	24
$N_c$ estimé sur les zones de vérité de terrain	8	21	11
ACCR	<b>97.66%</b>	47.20%	45.60%

- ***Classification à partir des attributs de tri-occurrences***

Dans cette section, nous évaluons les attributs de tri-occurrences proposés. Ces attributs sont calculés sur des voisinages cubiques de taille (3x3x3). Nous avons suivi les mêmes étapes que celles présentées dans la section précédente. Les résultats de classification sont tout d'abord comparés aux attributs statistiques du premier ordre (4 attributs : moyenne, variance, skewness et kurtosis), puis aux attributs issus de la transformée en ondelettes (2 attributs : moyenne, énergie).



La Figure 35 présente les différents résultats et le Tableau 12 donne la matrice de confusion du meilleur résultat (tri-occurrences et moyenne).

Les différents résultats obtenus confirment les conclusions données en 5.2.3 :

- Les attributs de tri-occurrences permettent une meilleure caractérisation des pixels des images réelles,
- L’ajout de l’attribut de moyenne aux attributs de tri-occurrences améliore encore les résultats de classification,
- Le taux de classification à partir des attributs de tri-occurrences s’améliore de 1.62% par rapport à la classification en utilisant directement les signatures spectrales comme attributs.

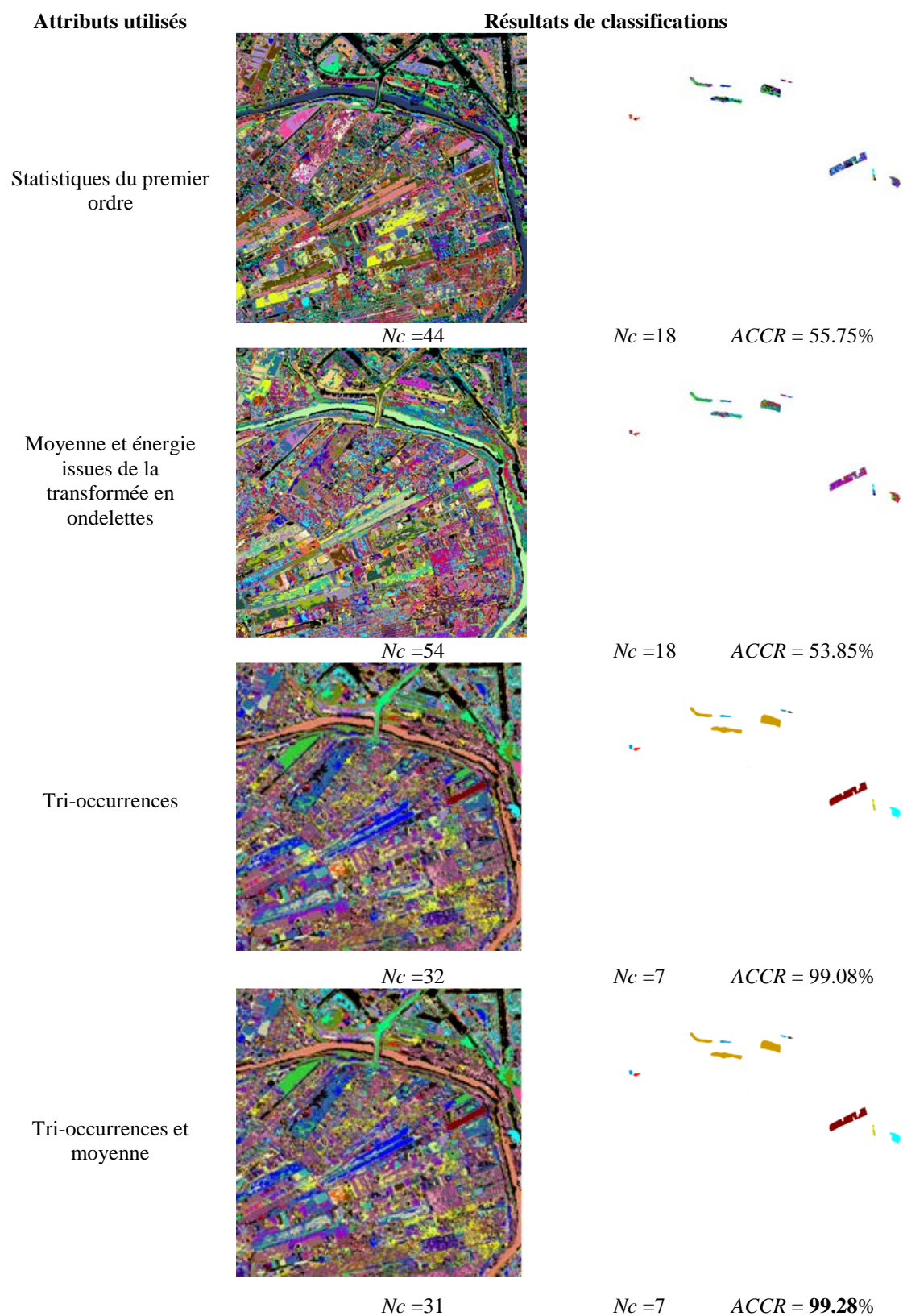


Figure 35 : Résultats de partitionnement de l'image hyperspectrale réelle avec différentes configurations d'attributs. 1ère colonne: partitionnements de l'image complète ; 2ème colonne : Résultats de détection de plantes invasives et non invasives sur les zones d'intérêt.

Tableau 12: Matrice de confusion correspondant aux résultats de partitionnement en 7 classes (attributs de tri-occurrence et moyenne). CCR en %, (.) : nombre de pixels.

Classes discriminées automatiquement par l'approche proposée (nombre de pixels)	Classes de la vérité de terrain (nombre de pixels)					
	<i>Phragmites australis</i> (556)	<i>Arundo donax</i> (4305)	<i>Tamarix</i> (162)	Pêcher (3115)	<i>Ulmus minor</i> (795)	<i>Pinus halepensis</i> (274)
<i>Phragmites australis</i> (532)	<b>95.68 %</b> (532)	0	0	0	0	0
<i>Arundo donax</i> (4305)	0	<b>100%</b> (4305)	0	0	0	0
<i>Tamarix</i> (162)	0	0	<b>100%</b> (162)	0	0	0
Pêcher (3115)	0	0	0	<b>100%</b> (3115)	0	0
<i>Ulmus minor</i> (795)	0	0	0	0	<b>100%</b> (795)	0
<i>Pinus halepensis</i> (274)	0	0	0	0	0	<b>100%</b> (274)
Classe anonyme 1 (24)	4.32% (24)	0	0	0	0	0

## 6.2 Identification du couvert paysager

### 6.2.1 Détection des pins

Dans le cadre de cette application, il s'agit d'évaluer le pouvoir discriminant de la chaîne d'analyse développée pour la détection d'une espèce de pins (cèdre) à partir de l'imagerie satellitaire multispectrale. L'image multispectrale utilisée ici a été acquise par le satellite d'observation Ikonos le 11 Juillet 2005, dans la région de Baabdat (Liban). Cette image est constituée de trois bandes (RVB) dont la taille du pixel est de 0,8 m. Conjointement à cette image, fournie par le Centre National de Télédétection du Liban dans le cadre d'un projet collaboratif CEDRE, des relevés de vérité de terrain partiels ont été communiqués pour permettre de calculer le taux de détection. La Figure 36 montre l'image originale (1000 x 1000 pixels), les masques des relevés de terrain fournis, ainsi que les différents résultats de partitionnement obtenus. Les conditions d'utilisation sont les mêmes que dans le cas de l'application précédente.

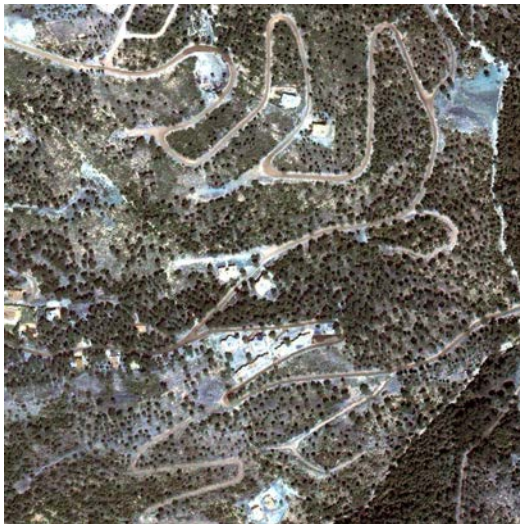
Pour l'ensemble des trois zones de la vérité de terrain, le Tableau 13 donne les taux d'occupation au sol par les pins pour 57 381 pixels considérés comme appartenant tous aux

pins. Dans certaine zones où le sol est pratiquement couvert par les pins (zone 1) le taux de détection estimé est de 67.21%. Par contre dans le cas des zones 2 et 3 où les pins sont dispersés, les taux estimés sont faibles si on se réfère aux observations de la vérité de terrain, avec des taux de 32.57 % et 30.38% respectivement. Toutefois, si l'on observe la densité des pins présents dans l'image, on peut considérer que ces taux correspondent plutôt à la proportion de surface occupée par les pins et non aux taux de détection. Le taux moyen d'occupation au sol par les pins et non le taux de détection pour l'ensemble des trois zones est de 40.78%.

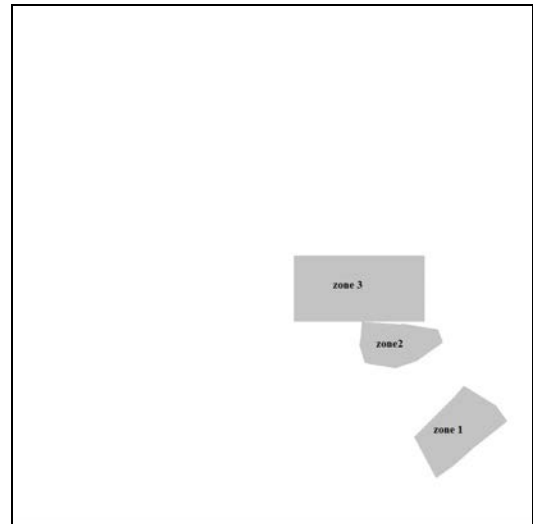
Cette application est un exemple typique qui démontre que l'utilisation de relevés de terrain approximatifs biaise fortement le taux de détection d'une méthode de classification. En effet, la vérité de terrain indique la présence de pins dans une zone, mais ne précise pas exactement la surface occupée par ces pins car l'accès à cette donnée est difficile.

Tableau 13: Taux d'occupation au sol des pins.

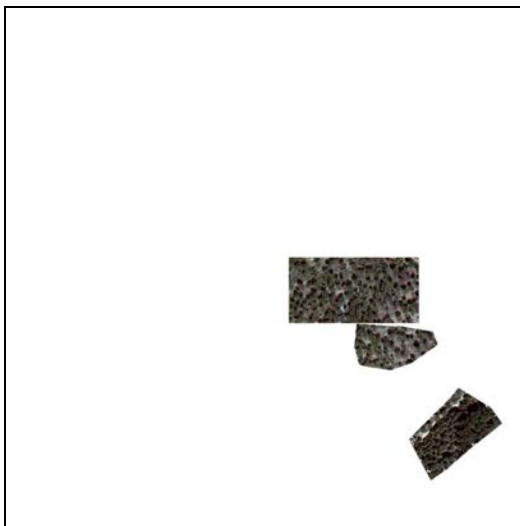
Zones	Nombre de pixels	Nombre de classes	Nombre de pixels détectés comme appartenant aux pins	Taux d'occupation au sol
Zone 1	15573	4	10467	67.21%
Zone 2	10683	4	3480	32.57%
Zone 3	31125	8	9458	30.38%



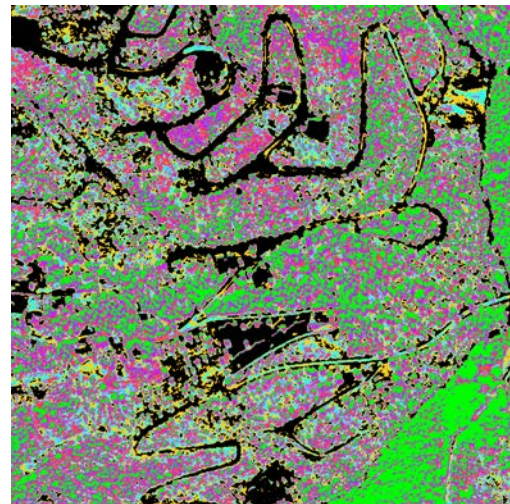
(a) : image originale (RVB) 1000x1000 pixels



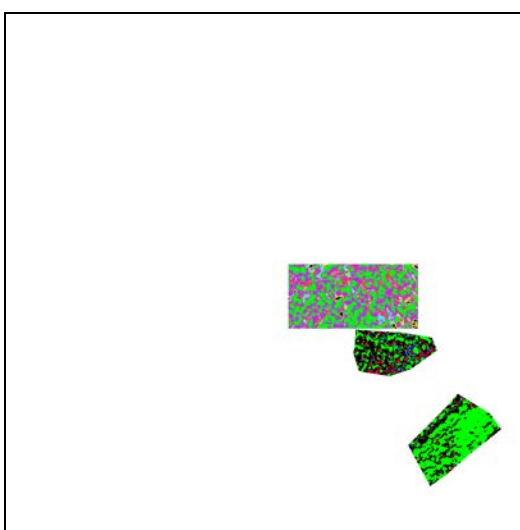
(b) : masque de la vérité de terrain



(c) : zones de l'image originale correspondant aux masques



(d) : image partitionnée complète



(e) : zones de la vérité de terrain partitionnées



(f) : Détection des pins

Figure 36 : Détection des pins par l'approche développée.

### 6.2.1 Identification des structures au sol

Cette troisième application concerne l'identification de quatre classes : bâtiment, sol nu, végétation, et plan d'eau, à partir d'une autre image satellitaire IKONOS de taille 154 x 152 pixels avec trois bandes spectrales et une résolution spatiale de 1 mètre. Des relevés ponctuels de terrain ont été réalisés à l'aide d'un système GPS pour quatre classes (bâtiment, sol nu, végétation, plan d'eau) représentant respectivement 18, 34, 26 et 52 points. En appliquant la chaîne d'analyse et de partitionnement, les taux de classification par classe en tenant compte du positionnement des points GPS et des pixels des 4 classes estimées, sont de 94.44% pour la classe bâtiment, 94.12% pour le sol nu et 88.46% pour la classe végétation et 76.92% pour la classe eau. Le taux de classification moyen (ACCR) est de 88.48%. Ce résultat est obtenu après réduction du nombre de pixels en considérant toute l'image, et il est très similaire à celui obtenu après découpage en quatre blocs. Les attributs utilisés sont les mêmes que ceux des applications précédentes.

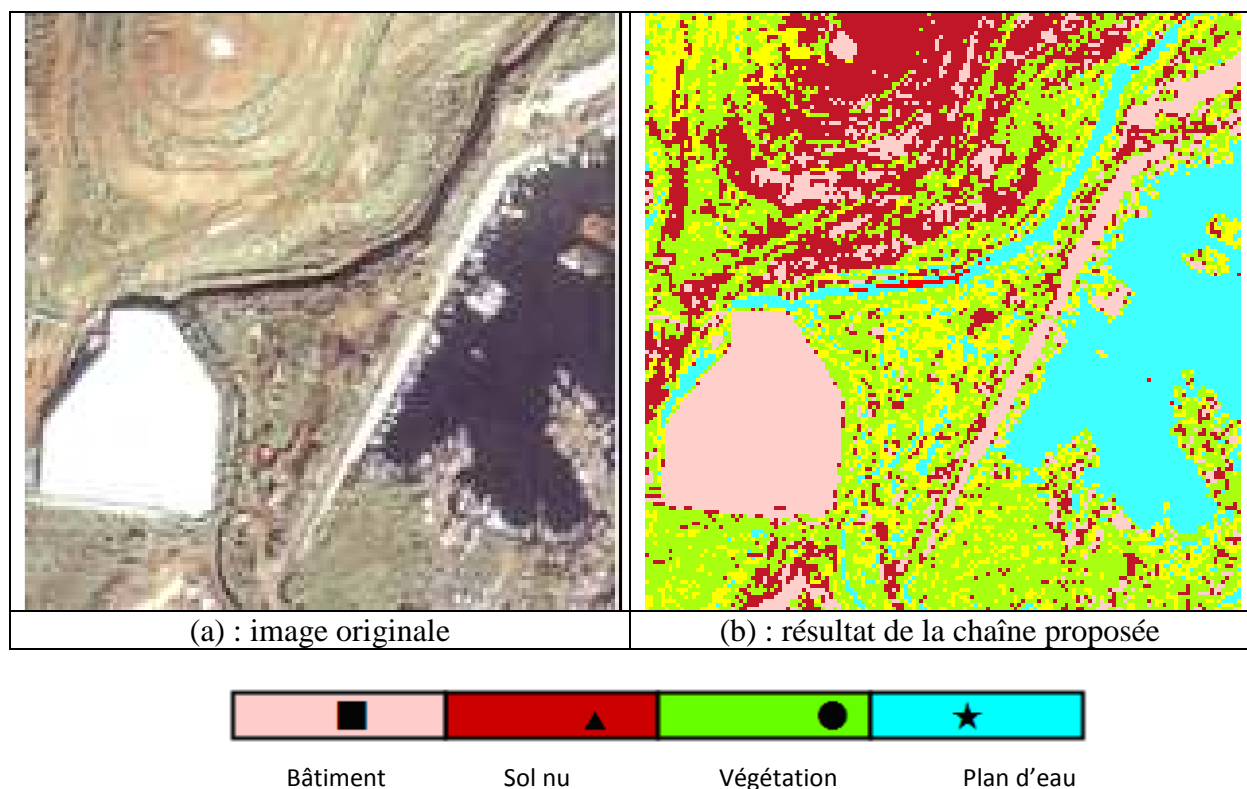


Figure 37 : Identification de quatre principales classes à partir d'Image multispectrale IKONOS.

## 6.4. Conclusion

L'approche de classification non paramétrique et non supervisée basée sur l'AP que nous avons développé dans ce travail a été validée sur une application réelle en identification d'espèces végétales. Les résultats ont montré la robustesse et la stabilité de cette méthode, qui exploite conjointement les attributs spectraux et spatiaux de tri-occurrences et de moyenne. L'application de l'AP après une étape préliminaire de réduction du nombre d'individus sur l'image hyperspectrale de grande dimension spatiale a été validée. Les résultats de la recherche automatique de la valeur optimale du paramètre  $p$  pour estimer le nombre de classes sont probants. Enfin, l'approche proposée donne de meilleurs résultats de classification par rapport aux approches classiques de classification semi-et non supervisées (C-moyennes, ISODATA, et SOM).



## Conclusion générale

Dans ce mémoire, nous avons proposé une approche originale de classification non supervisée dont la propriété fondamentale est de pouvoir s'appliquer aux données hyperspectrales de grande dimension spatiale. Pour atteindre cet objectif nous avons dû apporter une solution aux problèmes de caractérisation des pixels des images hyperspectrales, puis concevoir une méthode de réduction du nombre d'individus à classer qui permet l'application de la méthode de propagation d'affinité sans dégrader la qualité de la classification, et enfin développer une méthode d'estimation du nombre de classes optimal par propagation d'affinité, basée sur un critère d'évaluation de partitionnement.

Les conclusions majeures que l'on peut tirer de cette étude sont les suivantes :

- La pré-hiérarchisation des méthodes de caractérisation et de partitionnement des pixels a permis de dégager les principales aptitudes des méthodes proposées dans chacune de ces phases de traitement afin de mieux les exploiter.  
Cette étape a aussi permis de mettre en avant les meilleures méthodes en précisant leurs avantages et leurs limites.
- La méthode de caractérisation basée sur le calcul d'attributs de tri-occurrences spatiale-spectrale développée a montré de meilleures performances en caractérisation des pixels par rapport aux attributs classiques de la littérature (moments statistiques et issus de la transformée en ondelettes).
- La méthode de réduction du nombre d'individus à classer permet d'améliorer le taux de bonne classification et de résoudre le problème d'application de la méthode de classification par propagation d'affinité sur des données de grande dimension spatiale.
- La méthode d'estimation du nombre de classes basée sur la propagation d'affinité permet de résoudre le problème de cette dernière, lié au sur-partitionnement de l'image.

L'approche développée synthétisant ces différents points a été appliquée avec succès sur des données monocomposantes et hyperspectrales :



- L'application de l'approche proposée sur des données monocomposantes (images synthétiques) a montré que, en premier lieu, les attributs de tri-occurrences permettent une meilleure caractérisation des pixels des images. D'autre part, l'ajout de l'attribut moyenne améliore notablement le taux de bonne classification. En second lieu, la méthode classique de propagation d'affinité offre de meilleures performances en termes de classification par rapport aux méthodes classiques de partitionnement (C-moyennes, FCM et ISODATA) qui, rappelons le, sont des méthodes semi-supervisées nécessitant a minima la donnée du nombre de classes.
- L'application de l'approche proposée sur une application concrète en identification d'espèces végétales sur des données hyperspectrales réelles de grande dimension spatiale affirme les conclusions antérieures liées au pouvoir discriminant des attributs proposés et à la bonne estimation du nombre de classes par la méthode de propagation d'affinité modifiée.

Les quelques perspectives de ces travaux portent à la fois sur l'étape d'extraction d'attributs et sur l'étape de classification.

En effet, l'intégration de l'information spatiale et spectrale est un processus de grande complexité calculatoire qui peut encore être optimisé. Une perspective dans ce sens serait l'introduction d'une phase préliminaire de réduction de bandes permettant de préserver au maximum l'information spectrale.

L'introduction d'autres attributs d'ordre supérieur, par exemple les quadri-occurrences reste également une piste à explorer.

Quant à l'approche de classification par AP, le paramètre de préférence pourrait être estimé au cours d'une étape préliminaire afin d'éviter une recherche dichotomique nécessitant plusieurs exécutions de la méthode AP.

## Annexe A :

### Attributs spatiaux pour images monocomposantes

- **Attributs de cooccurrence :**

Les attributs calculés à partir de la matrice de cooccurrence  $M$  pour une distance  $d$  et une orientation  $\theta$  données avec un nombre de niveaux de gris  $N_\ell$  (le terme  $M_{d,\theta}(\ell_1, \ell_2)$  avec  $\ell_1$  et  $\ell_2$  deux niveaux de gris sont :

1. Second moment angulaire : ce paramètre mesure l'homogénéité de l'image; il est autant plus grand que l'image a des transitions de niveaux de gris dominantes

$$ASM = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M^2(\ell_1, \ell_2)$$

2. Contraste : il mesure le contraste d'une image; il est élevé lorsque l'on passe d'un pixel très clair à un pixel très foncé ou inversement.

$$CONT = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} (\ell_1 - \ell_2)^2 M(\ell_1, \ell_2)$$

3. Corrélation : il décrit la corrélation des niveaux de gris d'un pixel avec un autre pixel distant de  $d$  dans la direction  $\theta$ .

$$Corr = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \frac{(\ell_1 - F_i)(\ell_2 - F_j)}{\sigma_i \sigma_j} M(\ell_1, \ell_2)$$
$$F_i = \sum_{\ell_1=0}^{N_\ell-1} \ell_1 \times M_i(\ell_1) \text{ et } F_j = \sum_{\ell_2=0}^{N_\ell-1} \ell_2 \times M_j(\ell_2)$$

4. Variance : cet indicateur décrit la dispersion des transitions entre niveaux de gris.

$$VAR = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \left[ \ell_1 + 1 - \sum_{k=0}^{N_\ell-1} \sum_{l=0}^{N_\ell-1} k M(k, l) \right]^2 M(\ell_1, \ell_2)$$

5. Moment des différences inverses : ce paramètre renseigne sur l'importance des transitions entre niveaux de gris proches.

$$IDM = \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} \frac{M(\ell_1, \ell_2)}{1 + (\ell_1 - \ell_2)^2}$$

6. Moyennes des sommes :

$$SA = \sum_{h=2}^{2N_\ell} h \times (M_i M_j(h))$$

$$\text{avec : } M_i M_j(h) = \sum \sum_{\ell_1 + \ell_2 + 2 = h} M(\ell_1, \ell_2)$$

7. Variances des sommes :

$$SVAR = \sum_{h=2}^{2N_\ell} (h - SA)^2 M_i M_j(h)$$

8. Entropie des sommes :

$$Sentropie = - \sum_{h=2}^{2N_\ell} M_i M_j(h) \log(M_i M_j(h) + \varepsilon)$$

9. Entropie :

$$Entropie = - \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M(\ell_1, \ell_2) \log(M(\ell_1, \ell_2))$$

10. Variance des différences :

$$DVAR = \frac{N_\ell^2 \sum_{h=0}^{N_\ell-1} M_{ij}(h)^2 - \left( \sum_{l=0}^{N_\ell-1} M_{ij}(l) \right)^2}{N_\ell^4}$$

$$\text{Où } M_{ij}(h) = \sum_{\ell_1, \ell_2, |\ell_1 - \ell_2| = h} M(\ell_1, \ell_2)$$

11. Entropie de la différence :

$$DENTROPY = \sum_{h=0}^{N_\ell-1} M_{ij}(h) \log(M_{ij}(h) + \varepsilon)$$

12. Première information sur la mesure de corrélation :

$$ICORR1 = \frac{Hij - Hij'}{\max(Hi, Hj)}$$

$$\text{Où } Hij = - \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M(\ell_1, \ell_2) \log(M(\ell_1, \ell_2) + \varepsilon)$$

$$Hij' = - \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M(\ell_1, \ell_2) \log(M_i(\ell_1) + M_j(\ell_1) + \varepsilon)$$

$$Hi = - \sum_{\ell_1=0}^{N_\ell-1} M_i(\ell_1) \log(M_i(\ell_1) + \varepsilon), Hj = - \sum_{\ell_2=0}^{N_\ell-1} M_j(\ell_2) \log(M_j(\ell_2) + \varepsilon),$$

$$M_i(\ell_1) = \sum_{\ell_2=0}^{N_\ell-1} M(\ell_1, \ell_2)$$

$$\text{Et } M_i(\ell_2) = \sum_{\ell_1=0}^{N_\ell-1} M(\ell_1, \ell_2)$$

13. Deuxième information sur la mesure de corrélation :

$$ICORR2 = \sqrt{1 - e^{-2(Hij2 - Hij)}}$$

$$\text{Où } Hij2 = - \sum_{\ell_1=0}^{N_\ell-1} \sum_{\ell_2=0}^{N_\ell-1} M_i(\ell_1) M_j(\ell_2) \log(M_i(\ell_1) M_j(\ell_2) + \varepsilon)$$

14. Uniforme : cet indicateur est révélateur de la présence de texture.

$$Uniforme = \sum_{\ell_1=0}^{N_\ell-1} M(\ell_1, \ell_1)$$

15. Information contour : ce paramètre fournit une indication sur la présence de contours. Une valeur élevée traduit une transition prépondérante d'un niveau de gris avec les autres.

$$Icontour = \max_{\ell_1=0, N_\ell-1} \sum_{\ell_2=0}^{\ell_1} M(\ell_1, \ell_2)$$

- **Attributs de longueur de plages :**

Nous détaillons ici les attributs calculés à partir de la matrice de longueur de plages  $PL$  pour une distance et une orientation données.

Soient  $PL$  de taille  $NG \times L$ ,  $NG$  est le nombre de niveaux de gris de l'image et  $L$  la longueur maximale d'une plage.

1. importance des courtes plages :

$$SRE = \sum_{\ell=1}^{N_\ell} \sum_{l=1}^L \frac{PL(\ell, l, \theta)}{l^2}$$

2. importance des longues plages : cet indicateur est d'autant plus élevé que la probabilité de rencontrer des longues plages est élevée.

$$LRE = \sum_{\ell=1}^{N_\ell} \sum_{l=1}^L PL(\ell, l, \theta) l^2$$

3. distribution des niveaux de gris : ce paramètre est peu élevé pour une répartition hétérogène des plages pour plusieurs niveaux de gris.

$$GLD = \sum_{\ell=1}^{N_{\ell}} \left( \sum_{l=1}^L PL(\ell, l, \theta)^2 \right)$$

4. distribution des plages : cet attribut mesure la non-uniformité de la répartition des longueurs de plages indépendamment de leurs niveaux de gris.

$$RLD = \sum_{l=1}^L \left( \sum_{\ell=1}^{N_{\ell}} PL(\ell, l, \theta)^2 \right)$$

5. pourcentage de plages : ce paramètre mesure le nombre de plages par unité de surface, il est d'autant plus faible que les plages sont longues.

$$RLP = \sum_{\ell=1}^{NG} \sum_{l=1}^{L_{\max}} \frac{PL(\ell, l, \theta)}{N_v}$$

où  $n_t$  est le nombre de pixels de la fenêtre d'analyse.

## Annexe B :

### Critères d'évaluation comparés

Indice	Notation	Valeur optimale
Root-mean-square [130]	RMSSTD	minimum
R-squared [130]	RS	maximum
Indice de Calinski-Harabasz [131]	CH	maximum
Indice de Davies-Bouldin [132]	DB	minimum
Indice de Xie-Beni [133]	XIB	minimum
Indice de validité SD [134]	SD	minimum
Levine et Nazif [140]	Inter-LN	maximum
Indice de Silhouette [136]	Sil	maximum
Indice de Dunn [137]	Dunn	maximum

# Bibliographie

- [1] M. Hauta-Kasari, J. Parkkinen, T. Jaaskelainen, and R. Lenz, “Generalized co-occurrence matrix for multispectral texture analysis,” *Proceedings of the 13th International Conference on Pattern Recognition*, vol.2, pp. 785–789, 1996.
- [2] V. Arvis, C. Debain, M. Berducat, and A. Benassi, “Generalization of the cooccurrence matrix for colour images: application to colour texture classification,” *Image Analysis & Stereology*, vol. 23, no. 1, pp. 63–72, May 2011.
- [3] F. Tsai, C.-K. Chang, J.-Y. Rau, T.-H. Lin, and G.-R. Liu, “3D Computation of Gray Level Co-occurrence in Hyperspectral Image Cubes,” in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 429–440, 2007.
- [4] R. Khelifi, M. Adel, and S. Bourennane, “Multispectral texture characterization: application to computer aided diagnosis on prostatic tissue images,” *EURASIP Journal on Advances Signal Processing*, vol. 2012, no. 1, pp. 1–13, Dec. 2012.
- [5] B. J. Frey and D. Dueck, “Clustering by Passing Messages Between Data Points,” *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
- [6] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967.
- [7] P. J. F. Groenen and K. Jajuga, “Fuzzy clustering with squared Minkowski distances,” *Fuzzy Sets and Systems*, vol. 120, no. 2, pp. 227–237, Jun. 2001.
- [8] G. H. Ball and D. J. Hall, “A clustering technique for summarizing multivariate data,” *Syst. Res.*, vol. 12, no. 2, pp. 153–155, Mar. 1967.
- [9] S. Jia, Y. Qian, and Z. Ji, “Band Selection for Hyperspectral Imagery Using Affinity Propagation,” *Digital Image Computing: Techniques and Applications (DICTA)*, pp. 137–141, 2008.
- [10] S. Jia, Z. Ji, Y. Qian, and L. Shen, “Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 531–543, Apr. 2012.

- [11] M. Soltani, K. Chehdi, and C. Cariou, "Affinity propagation for large size hyperspectral image classification," in *Proceedings of the SPIE*, vol. 8892, p. 88920P–88920P–10, 2013.
- [12] K. Chehdi, M. Soltani, and C. Cariou, "Pixel classification of large-size hyperspectral images by affinity propagation," *Journal of Applied Remote Sensing*, vol. 8, no. 1, pp. 083567–083567, 2014.
- [13] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *International Journal of Remote Sensing*, vol. 17, no. 9, pp. 1623–1632, Jun. 1996.
- [14] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. John Wiley & Sons, 2007.
- [15] J. J. Workman, "Review of Process and Non-invasive Near-Infrared and Infrared Spectroscopy: 1993–1999," *Applied Spectroscopy Reviews*, vol. 34, no. 1–2, pp. 1–89, Jul. 1999.
- [16] C. P. Warren, D. Even, W. Pfister, K. Nakanishi, A. Velasco, D. Breitwieser, S. Yee, and J. Naungayan, "Miniaturized visible near-infrared hyperspectral imager for remote-sensing applications," *Optical Engineering*, vol. 51, no. 11, pp. 111720–1, Aug. 2012.
- [17] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, Jan. 2014.
- [18] Q. Li, W. Wang, C. Ma, and Z. Zhu, "Detection of physical defects in solar cells by hyperspectral imaging technology," *Optics & Laser Technology*, vol. 42, no. 6, pp. 1010–1013, Sep. 2010.
- [19] V. Farley, M. Chamberland, P. Lagueux, A. Vallières, A. Villemare, and J. Giroux, "Chemical agent detection and identification with a hyperspectral imaging infrared sensor," *Proc. SPIE 6661, Imaging Spectrometry XII*, p. 66610L–66610L–9, 2007.
- [20] G. Elmasry, M. Kamruzzaman, D.-W. Sun, and P. Allen, "Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review," *Critical Reviews in Food Science and Nutrition*, vol. 52, no. 11, pp. 999–1023, Nov. 2012.
- [21] L. M. Dale, A. Thewis, C. Boudry, I. Rotar, P. Dardenne, V. Baeten, and J. A. F. Pierna, "Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review," *Applied Spectroscopy Reviews*, vol. 48, no. 2, pp. 142–159, Mar. 2013.



- [22] F. Vagni, "Survey of Hyperspectral and Multispectral Imaging Technologies," RTO/NATO Technical Report TR-SET-065-P3, 2007.
- [23] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [24] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, Sep. 1999.
- [25] S. Tadjudin and D. Landgrebe, "Classification of high dimensional data with limited training samples," *ECE Technical Reports*, Apr. 1998.
- [26] R. Zwiggelaar, "A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops," *Crop Protection*, vol. 17, no. 3, pp. 189–206, May 1998.
- [27] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
- [28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [29] C. M. Bishop, *Neural networks for pattern recognition*. Oxford : New York: Clarendon Press ; Oxford University Press, 1995.
- [30] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines: and other kernel-based learning methods*. Cambridge ; New York: Cambridge University Press, 2000.
- [31] P. Chaovalit and L. Zhou, "Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches," *Proceedings of the 38th Hawaii International Conference on System Sciences*.pp. 1-9, 2005.
- [32] J. Morlet, "Sampling Theory and Wave Propagation," *Issues in Acoustic Signal — Image Processing and Recognition*, Springer Berlin Heidelberg, pp. 233–261, 1983.
- [33] M.M. Galloway, "Texture analysis using gray level run lengths" *Graphical Models And Image Processing*, vol. 4, pp. 172-179, 1975.
- [34] G. Casella and R. L. Berger, *Statistical Inference*, 2nd edition. Australia ; Pacific Grove, CA: Cengage Learning, 2001.
- [35] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

- [36] L. Soh and C. Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery using Gray Level Co-occurrence Matrices," *IEEE Transactions On Geoscience And Remote Sensing*, pp. 780–795, 1999.
- [37] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, Feb. 2002.
- [38] A. G.-D. Carlos Aviles-Cruz, "Utilisation des statistiques d'ordre supérieur pour l'analyse d'images texturées," *GRETSI 16° Colloque sur le traitement du signal et des images*, pp. 395-398, 1997.
- [39] C. Rosenberger, *Mise en oeuvre d'un système adaptatif de segmentation d'images*, Thèse de Doctorat, Université de Rennes 1, 1999.
- [40] N. Voisine, *Approche adaptative de coopération hiérarchique de méthodes de segmentation : application aux images multicomposantes*, Thèse de Doctorat, Université de Rennes 1, 2002.
- [41] C. Rosenberger and K. Chehdi, "Toward a complete adaptive analysis of an image," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 292–298, 2003.
- [42] P. Brodatz, *Textures : A Photographic Album for Artists and Designers*, Dover, New-York. 1966.
- [43] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [44] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
- [45] J. P. Cocquerez and S. Philipp-Foliguet, *Analyse d'images: filtrage et segmentation*. MASSON, 1995.
- [46] S. Arivazhagan and L. Ganesan, "Texture segmentation using wavelet transform," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3197–3203, Dec. 2003.
- [47] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1513–1521, Jun. 2003.
- [48] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

- [49] T. Aydin, Y. Yemez, E. Anarim, and B. Sankur, "Multidirectional and multiscale edge detection via M-band wavelet transform," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1370–1377, Sep. 1996.
- [50] C. S. Lu, P. C. Chung, and C. F. Chen, "Unsupervised texture segmentation via wavelet transform," *Pattern Recognition*, vol. 30, no. 5, pp. 729–742, May 1997.
- [51] Y. Meyer, *Ondelettes, filtres miroirs en quadrature et traitement numérique de l'image*. Springer, 1990.
- [52] F. Truchetet, *Ondelettes pour le signal numérique*. Paris: Hermès, 1998.
- [53] A. Mojsilovic, M. V. Popovic, and D. M. Rackov, "On the selection of an optimal wavelet basis for texture characterization," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2043–2050, Dec. 2000.
- [54] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186–1191, Nov. 1993.
- [55] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 429–441, Oct. 1993.
- [56] C. F. V. Navid Fatemi-ghomi, "*Performance measures for Wavelet-based Segmentation Algorithms*," Surrey University, Thèse de doctorat, 1997.
- [57] S. Livens, P. Scheunders, G. V. de Wouwer, D. V. Dyck, H. Smets, J. Winkelmans, and W. Bogaerts, "A Texture Analysis Approach to Corrosion Image Classification," *Microscopy, Microanalysis, Microstructures*, vol. 7, pp. 143–152, 1996.
- [58] M. Simard, S. S. Saatchi, and G. De Grandi, "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2310–2321, Sep. 2000.
- [59] A. Gagalowicz, "A new method for texture fields synthesis: some applications to the study of human vision.," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 3, no. 5, pp. 520–533, May 1981.
- [60] A. Gagalowicz, *Vers un modèle de textures*, Thèse de Doctorat, Université Pierre et Marie Curie, 1983.
- [61] A. B. P. Reuze, "Caractérisation d'images texturées basée sur les statistiques d'ordre trois," *15<sup>e</sup> Colloque Gretsi*, pp. 645-648, 1995.
- [62] C. Coroyer, *Apport des corrélations d'ordre élevé à l'analyse de textures non-gaussiennes*, *Colloque Gretsi*, pp. 129-138, 1998.

- [63] H. R. Kalluri, S. Prasad, and L. M. Bruce, "Decision-Level Fusion of Spectral Reflectance and Derivative Information for Robust Hyperspectral Land Cover Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4047–4058, Nov. 2010.
- [64] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
- [65] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer, 2003.
- [66] C. Münzenmayer, H. Volk, C. Küblbeck, K. Spinnler, and T. Wittenberg, "Multispectral Texture Analysis Using Interplane Sum- and Difference-Histograms," in *Pattern Recognition*, L. V. Gool, Ed. Springer Berlin Heidelberg, pp. 42–49, 2002.
- [67] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [68] R. Kondepudy and G. Healey, "Use of invariants for recognition of three-dimensional color textures," *Journal of the Optical Society of America*, vol. 11, no. 11, pp. 3037–3049, Nov. 1994.
- [69] O. Rajadell, P. García-Sevilla, and F. Pla, "Textural Features for Hyperspectral Pixel Classification," in *Pattern Recognition and Image Analysis*, pp. 208–216, 2009.
- [70] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, "Texture feature analysis using a gauss-Markov model in hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.
- [71] S. Sarkar and G. Healey, "Hyperspectral texture classification using generalized Markov fields," *Optical Engineering*, vol. 43, no. 12, pp. 3038–3044, 2004.
- [72] L. Lepisto, I. Kunttu, J. Autio, and A. Visa, "Classification method for colored natural textures using Gabor filtering," in *12th International Conference on Image Analysis and Processing*, pp. 397–401, 2003.
- [73] M. Hauta-Kasari, J. Parkkinen, T. Jaaskelainen, and R. Lenz, "Generalized co-occurrence matrix for multispectral texture analysis," in *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 2, pp. 785–789, 1996.
- [74] R. Khelifi, M. Adel, and S. Bourennane, "Texture classification for multi-spectral images using spatial and spectral Gray Level Differences," in *2nd International*

- Conference on Image Processing Theory Tools and Applications (IPTA)*, pp. 330–333, 2010.
- [75] R. Khelifi, M. Adel, S. Bourennane, and A. Moussaoui, “Generalized gray level dependence method for prostate cancer classification,” in *7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, pp. 295–298, 2011.
  - [76] L. Lepistö, J. Kunttu, J. Autio, and A. Visa, “Rock Image Classification Using Non-Homogenous Textures and Spectral Imaging,” in *Spectral Imaging, WSCG SHORT PAPERS proceedings, WSCG’2003, Plzen, Czech Republic*, 2003.
  - [77] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, “Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
  - [78] J. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, “Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis,” in *IEEE International Proceedings of Geoscience and Remote Sensing Symposium, IGARSS*, vol. 1, pp. 4–25, 2005.
  - [79] A. Plaza, P. Martinez, R. Perez, and J. Plaza, “Spatial/spectral endmember extraction by multidimensional morphological operations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.
  - [80] D. Mahmoud-Ghoneim, G. Toussaint, J. M. Constans, and J. D. de Certaines, “Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas,” *Magnetic Resonance Imaging* ol. 21, no. 9, pp. 983–987, Nov. 2003.
  - [81] R. Khelifi, M. Adel, and S. Bourennane, “Spatial and spectral dependance co-occurrence method for multi-spectral image texture classification,” in *17th IEEE International Conference on Image Processing (ICIP)*, pp. 4361–4364, 2010.
  - [82] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
  - [83] P. T. Kohonen, “Self-Organizing Feature Maps,” in *Self-Organization and Associative Memory*, Springer Berlin Heidelberg, pp. 119–157, 1989.
  - [84] Y. Linde, A. Buzo, and R. M. Gray, “An Algorithm for Vector Quantizer Design,” *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, Jan. 1980.
  - [85] R. Krishnapuram and J. M. Keller, “A possibilistic approach to clustering,” *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, May 1993.

- [86] D. S. Xia Lin, "A Self-Organizing Semantic Map for Information Retrieval," in *Proceedings Of The 14th Annual International ACM SIGIR Conference On Research And Development In Information Retrieval*, pp. 262–269, 1991.
- [87] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*. CRC Press, 2009.
- [88] Q. Wen and M. E. Celebi, "Hard versus fuzzy c-means clustering for color quantization," *EURASIP Journal Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–12, Dec. 2011.
- [89] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, Jan. 1979.
- [90] K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," *IEEE Transactions On Pattern Analysis And Machine Intelligence* vol. 24, no. 7, pp. 881–892, 2002.
- [91] C. Elkan, *Using the Triangle Inequality to Accelerate k-Means*. in *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 147–153, 2003.
- [92] M. Song and S. Rajasekaran, "Fast k-Means Algorithms with Constant Approximation," in *Algorithms and Computation*, Springer Berlin Heidelberg, pp. 1029–1038, 2005.
- [93] G. Karypis, E.-H. Han, and V. Kumar, "Chameleon: hierarchical clustering using dynamic modeling," *Computer*, vol. 32, no. 8, pp. 68–75, Aug. 1999.
- [94] M. Kunt, *Reconnaissance des formes et analyse de scènes*. Collection Electricité, 2000.
- [95] E. L. J. Bohez, "Two level cluster analysis based on fractal dimension and iterated function systems (IFS) for speech signal recognition," *Electronics Letters*, vol. 28, no. 15, pp. 291 – 294, 1998.
- [96] M. F. Hussin, M. S. Kamel, and M. H. Nagi, "An Efficient Two-Level SOMART Document Clustering Through Dimensionality Reduction," in *Neural Information Processing*, Springer Berlin Heidelberg, pp. 158–165, 2004.
- [97] Y. B. Sébastien Guérif, "Selection of clusters number and features subset during a two-levels clustering task," in *Proceeding of Artificial Intelligence and Soft Computing*, pp. 28–33, 2006.
- [98] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.
- [99] M. Mitchell, *An Introduction to Genetic Algorithms*. MIT Press, 1998.

- [100] D. Pham and D. Karaboga, *Intelligent Optimisation Techniques - Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*, Sep. 2010
- [101] D. F. Rothlauf, “Representations for Genetic and Evolutionary Algorithms,” in *Representations for Genetic and Evolutionary Algorithms*, Springer Berlin Heidelberg, pp. 9–32, 2006.
- [102] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.*, University of Michigan, Ann Arbor, MI, USA, 1975.
- [103] S. Bandyopadhyay and U. Maulik, “Nonparametric genetic clustering: comparison of validity indices,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 31, no. 1, pp. 120–125, Feb. 2001.
- [104] T. J. Ross, *Fuzzy Logic with Engineering Applications*. John Wiley & Sons, 2009.
- [105] S. Bandyopadhyay and U. Maulik, “Genetic clustering for automatic evolution of clusters and application to image classification,” *Pattern Recognition*, vol. 35, no. 6, pp. 1197–1208, Jun. 2002.
- [106] C.-C. Hung, S. Kulkarni, and B.-C. Kuo, “A New Weighted Fuzzy C-Means Clustering Algorithm for Remotely Sensed Image Classification,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 543–553, Jun. 2011.
- [107] F. Yao and Y. Qian, “Band selection based gaussian processes for hyperspectral remote sensing images classification,” in *16th IEEE International Conference on Image Processing (ICIP)*, pp. 2845–2848, 2009.
- [108] Y. Yang, H.-Y. Ha, F. C. Fleites, and S.-C. Chen, “A Multimedia Semantic Retrieval Mobile System Based on HCFGs,” *IEEE MultiMedia*, vol. 21, no. 1, pp. 36–46, Jan. 2014.
- [109] H. Xiao and P. Guo, “Iris Image Analysis Based on Affinity Propagation Algorithm,” *Advances in Neural Networks*, pp. 943–949, 2009.
- [110] C.-D. Wang, J.-H. Lai, C. Y. Suen, and J.-Y. Zhu, “Multi-Exemplar Affinity Propagation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2223–2237, Sep. 2013.
- [111] Z. Liu, P. Li, Y. Zheng, and M. Sun, “Clustering to Find Exemplar Terms for Keyphrase Extraction,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Vol. 1, pp. 257–266, 2009.
- [112] Z.-J. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang, “Visual Query Suggestion,” in *Proceedings of the 17th ACM International Conference on Multimedia*, pp. 15–24, 2009.

- [113] K. Lindorff-Larsen and J. Ferkinghoff-Borg, "Similarity measures for protein ensembles," *PloS one*, vol. 4, no. 1, pp. 15-22, 2009.
- [114] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Computer Vision*, vol. 3, no. 4, pp. 213–222, Dec. 2009.
- [115] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.
- [116] S. Jia, Y. Qian, J. Li, W. Liu, and Z. Ji, "Feature extraction and selection hybrid algorithm for hyperspectral imagery classification," in *IEEE International on Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 72–75, 2010.
- [117] C. Yang, S. Liu, L. Bruzzone, R. Guan, and P. Du, "A Feature-Metric-Based Affinity Propagation Technique for Feature Selection in Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1152–1156, Sep. 2013.
- [118] C. Yang, L. Bruzzone, F. Sun, L. Lu, R. Guan, and Y. Liang, "A Fuzzy-Statistics-Based Affinity Propagation Technique for Clustering in Multispectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2647–2659, Jun. 2010.
- [119] C. Yang, L. Bruzzone, R. Guan, L. Lu, and Y. Liang, "Incremental and Decremental Affinity Propagation for Semisupervised Clustering in Multispectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1666–1679, Mar. 2013.
- [120] I. Givoni, C. Chung, and B. J. Frey, "Hierarchical Affinity Propagation," in *Computer Science*, pp. 212–221, Feb. 2012.
- [121] R. Guan, X. Shi, M. Marchese, C. Yang, and Y. Liang, "Text Clustering with Seeds Affinity Propagation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 4, pp. 627–637, Apr. 2011.
- [122] J. A. Cardille, J. C. White, M. A. Wulder, and T. Holland, "Representative Landscapes in the Forested Area of Canada," *Environmental Management*, vol. 49, no. 1, pp. 163–173, Jan. 2012.
- [123] M. Sharma, "Performance evaluation of image segmentation and texture extraction methods in scene analysis," *Master of philosophy in computer Science to the university of Exter, Exter, UK*, 2001.



- [124] S. Ghandour, *Segmentation d'images couleurs par morphologie mathématique : application aux images microscopiques*, Thèse de doctorat, Université de Toulouse, 2010.
- [125] J.-P. Cocquerez and J. Devars, "Edge detection in aerial pictures: new operator," *Traitement signal*, vol. 2, no. 1, pp. 45-65, 1985.
- [126] X.-P. Zhang and M. D. Desai, "Segmentation of bright targets using wavelets and adaptive thresholding," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1020–1030, Jul. 2001.
- [127] L. Vinet, *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques segmentation and region matching of a stereo pair of images*, Thèse de Doctorat, Université Paris 09, 1991.
- [128] G. Cleuziou, *Une méthode de classification non-supervisée pour l'apprentissage de règles et la recherche d'information*, Thèse de Doctorat, Université d'Orléans, 2004.
- [129] L. Lebart, A. Morineau, J-P. Fénelon, *Traitement des données statistiques*, 2000.
- [130] S. Sharma, *Applied Multivariate Techniques*, Thèse de Doctorat, Université South Carolina, 1995.
- [131] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974.
- [132] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
- [133] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
- [134] M. Halkidi, M. Vazirgiannis, and Y. Batistakis, "Quality Scheme Assessment in the Clustering Process," in *Principles of Data Mining and Knowledge Discovery*, Springer Berlin Heidelberg, pp. 265–276, 2000.
- [135] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: finding the optimal partitioning of a data set," in *Proceedings IEEE International Conference on Data Mining*, pp. 187–194, 2001.
- [136] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.
- [137] J. C. Dunn, "Well-Separated Clusters and Optimal Fuzzy Partitions," *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, Jan. 1974.

- [138] M. Moghrani, *Segmentation coopérative et adaptative d'images multicomposantes : application aux images CASI*, Thèse de Doctorat, Université Rennes 1, 2007.
- [139] K. Chehdi. C. Kermad, "Segmentation d'images par multi-seuillage et fusion de régions labellisées minimisant un critère de similarité," *15<sup>e</sup> Colloque sur le traitement du signal et des images*, pp. 641-644, 1995.
- [140] M. D. Levine and A. M. Nazif, "Dynamic Measurement of Computer Generated Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 155–164, Mar. 1985.
- [141] D. Dasgupta, *Artificial Immune Systems and Their Applications*. Berlin, Germany: Springer-Verlag, 1999.
- [142] J. H. Carter, "The immune system as a model for pattern recognition and classification," *Journal Of Informatics In Health And Biomedicine*, vol. 7, no. 3, pp. 28–41, 2000.
- [143] J. Timmis, M. Neal, and J. E. Hunt, "An artificial immune system for data analysis," *Biosystem*, vol. 55, no. 1/3, pp. 143–150, 2000.
- [144] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.
- [145] Z. Zhu, S. Jia, and Z. Ji, "Towards a Memetic Feature Selection Paradigm," *IEEE Computational Intelligence Magazine*, vol. 5, no. 2, pp. 41–53, 2010.
- [146] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. 110–122, Sep. 2009.
- [147] A. Singla and M. Karambir, "Comparative Analysis & Evaluation of Euclidean Distance Function and Manhattan Distance Function Using K-means Algorithm," *International Journal of Advanced Research in Computer Science and Software Engineering (IJARSSE)*, vol. 2, no. 7, pp. 298–300, Juin. 2012.
- [148] G. Khosla, N. Rajpal, and J. Singh, "Evaluation of Euclidean and Manhatttan Metrics In Content Based Image Retrieval System" *Journal of Engineering Research and Applications*, vol. 4, no. 9, pp. 43–49, Sep 2014.

# Liste des tableaux

Tableau 1 : Caractéristiques de l'imagerie RVB, de l'imagerie MS, et de l'imagerie MS. ....	13
Tableau 2: Mise à jour des degrés d'appartenance dans la famille des C-moyennes.....	38
Tableau 3 : Correspondance entre le système immunitaire humain et UAIC .....	43
Tableau 4: Comparaison des méthodes de réduction des bandes des images hyperspectrales et classification avec SVM (ACCR : Taux de bonne classification moyen (en %))......	54
Tableau 5: Comparaison des méthodes de classification sur des images multispectrales (ACCR en %). ....	55
Tableau 6: Indices de similarité entre deux individus $x_i$ et $x_j$ .....	58
Tableau 7: Valeur optimale du paramètre de préférence, nombre de classes obtenues et ACCR correspondant pour différents jeux d'attributs et différents critères d'évaluation.....	76
Tableau 8: Résultats de classification en fonction de la taille des blocs .....	99
Tableau 9 : Matrice de confusion (5 classes) correspondant au résultat de partitionnement de la Figure 32(c), CCR en %, (.) : nombre de pixels.....	103
Tableau 10: Matrice de confusion correspondant aux résultats de partitionnement à 8 classes.....	108
Tableau 11: Comparaison des performances de trois classifieurs sur la zone d'image Cieza (1000x1000 pixels, 62 bandes spectrales).....	109
Tableau 12: Matrice de confusion correspondant aux résultats de partitionnement en 7 classes (attributs de tri-occurrence et moyenne). CCR en %, (.) : nombre de pixels.....	112
Tableau 13: Taux d'occupation au sol des pins .....	113

# Liste des figures

Figure 1: De la vision numérique monochrome à l'imagerie hyperspectrale [14] .....	13
Figure 2 : Exemple de décomposition en 3 niveaux d'ondelettes.....	24
Figure 3 : Évolution du voisinage en fonction de la résolution.....	26
Figure 4 : Représentation de deux voxels $x_i$ et $x_j$ pour une distance $d=1$ et angle $\theta=45^\circ$ , $\theta=90^\circ$ .....	29
Figure 5: Taux de bonne classification en fonction des différents attributs. ....	30
Figure 6: Représentation spectrale de deux pixels $x_i$ et $x_j$ pour une distance $d=1$ et $\theta=45^\circ$ .....	31
Figure 7: Taux de bonne classification en fonction des différents attributs et du nombre de bandes retenues. ....	32
Figure 8: Topologie de l'algorithme SOM.....	40
Figure 9: Procédures d'échanges de messages de l'AP.....	56
Figure 10: Procédé de construction d'une image synthétique.....	59
Figure 11: Exemple d'images synthétiques à partitionner .....	61
Figure 12: Résultats de partitionnement des images de la Figure 11 en fonction du choix de l'indice de similarité et des attributs.....	61
Figure 13: Taux de classification en fonction des indices de similarité .....	62
Figure 14: Taux de classification moyens pour les 3 images en fonction des indices de similarité.....	63
Figure 15: Image originale et image test construite .....	73
Figure 16: Les signatures spectrales de la vérité de terrain des cinq classes de l'image hyperspectrale reconstruite .....	73

Figure 17: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la Figure 15(c) (attribut : signature spectrale d'un pixel) .....	74
Figure 18: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la Figure 15(c) (attributs : statistiques du 1 <sup>er</sup> ordre calculés sur une fenêtre 3x3) .....	75
Figure 19: Évolution des critères d'évaluation en fonction du paramètre de préférence pour l'image hyperspectrale de la figure 15(c) (attributs : moyenne et énergie des coefficients d'ondelettes calculés sur une fenêtre 3x3). .....	75
Figure 20: Chaîne de partitionnement d'images .....	81
Figure 21: Les 28 combinaisons spatiales possibles .....	83
Figure 22: Résultats de partitionnement de trois images synthétiques avec différentes configurations d'attributs en utilisant l'AP standard .....	87
Figure 23: Performances des résultats de partitionnement des trois images synthétiques avec différentes configurations d'attributs. ....	88
Figure 24: Différentes dispositions des fenêtres d'analyses pour l'extraction d'attributs .....	90
Figure 25: Quelques exemples de configurations de tri-occurrences dans un cube 3x3x3 .....	90
Figure 26: Résultats de partitionnement de l'image hyperspectrale synthétique avec différentes configurations d'attributs .....	92
Figure 27: Réduction du nombre de pixels à classer par appariement en fonction de la taille des blocs par deux approches.....	98
Figure 28: Performances de l'approche de réduction et de classification en fonction de la taille des blocs de découpage de l'image originale.....	99
Figure 29: Performance de l'AP (ACCR (%)) en fonction du nombre de pixels après réduction par appariement (L1) et par type d'attributs.....	100
Figure 30 : Résultats de classification de l'AP (nombre de classes fixé à 5) avec et sans réduction .....	100
Figure 31 : Résultat de classification de l'AP standard en mode non supervisé sans et avec réduction .....	101
Figure 32 : Partitionnement de l'image hyperspectrale synthétique par l'AP avec étape de réduction du nombre d'individus : évolution du nombre de classes en fonction de la valeur du paramètre $p$ . ....	102
Figure 33 : Image hyperspectrale Cieza acquise par le capteur AISA Eagle et zones d'intérêt à partitionner....	105
Figure 34 : Partitionnement de l'image hyperspectrale Cieza en fonction de la valeur du paramètre $p$ : 1ère colonne: partitionnements de l'image complète ; 2ème colonne : Résultats de détection de plantes invasives et non invasives sur les zones d'intérêt.....	107
Figure 35 : Résultats de partitionnement de l'image hyperspectrale réelle avec différentes configurations d'attributs. 1ère colonne: partitionnements de l'image complète ; 2ème colonne : Résultats de détection de plantes invasives et non invasives sur les zones d'intérêt. ....	111
Figure 36 : Détection des pins par l'approche développée.....	114
Figure 37 : Identification de quatre principales classes à partir d'Image multispectrale IKONOS.....	115